

The Training on Local Language Computing Fundamentals

Internationalization/Localization

ISO/IEC 10646 (Unicode)

Hints on Asian Input method

2004-1-19/23 Lahore, Pakistan

Takayuki K. Sato

Chief Researcher, CICC-Japan

Part-1 I18N/L10N

Key Messages

Localization is easy job,
but difficult business

i18n/l10n structure with
SINGLE BINARY is a way to go

Objective of Localization

Is NOT writing a program that is capable to I/O local character set.

Is to provide a tool to bridge
a digital divide

Is to make a country competitive

Bridging Digital Divide

- Use the world class software
- In local custom and language
- with interoperability within world network
- with the compatibility of data
- at the same price and availability

Reinvented equivalent products does not meet the objective

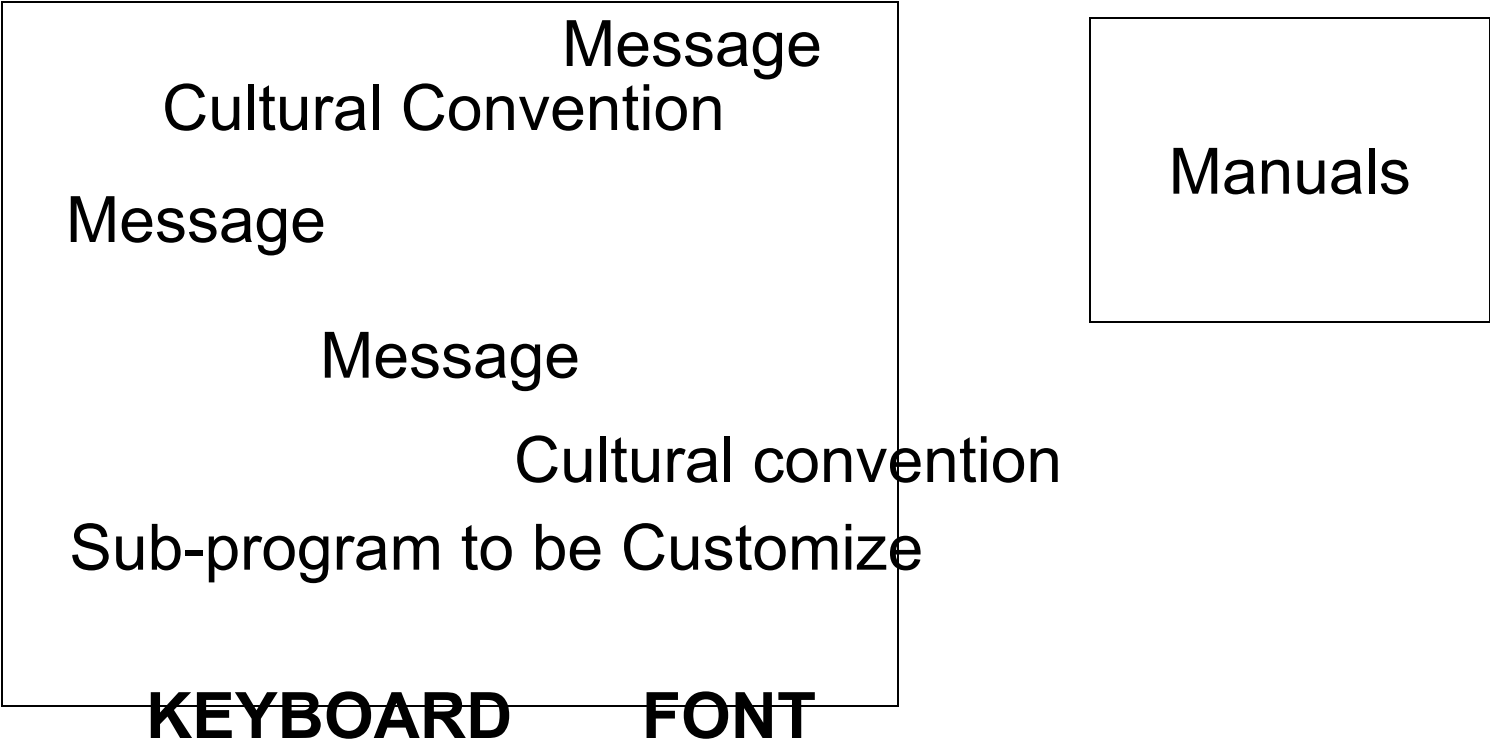
Local User Requests

following in local custom and language

- Operating Manuals
- Date Processing in local custom
- Out put in local character (character handling)
- Custom Functions
- Friendly Input
- Updates in Timely and free of charge
- Cultural Conventions
- User Interface

Meeting all above should be called **localization**

Old Style Hard Code Model



Costly and Time Consuming Localization

Hard Coded Model is “Search, Modify and Test”

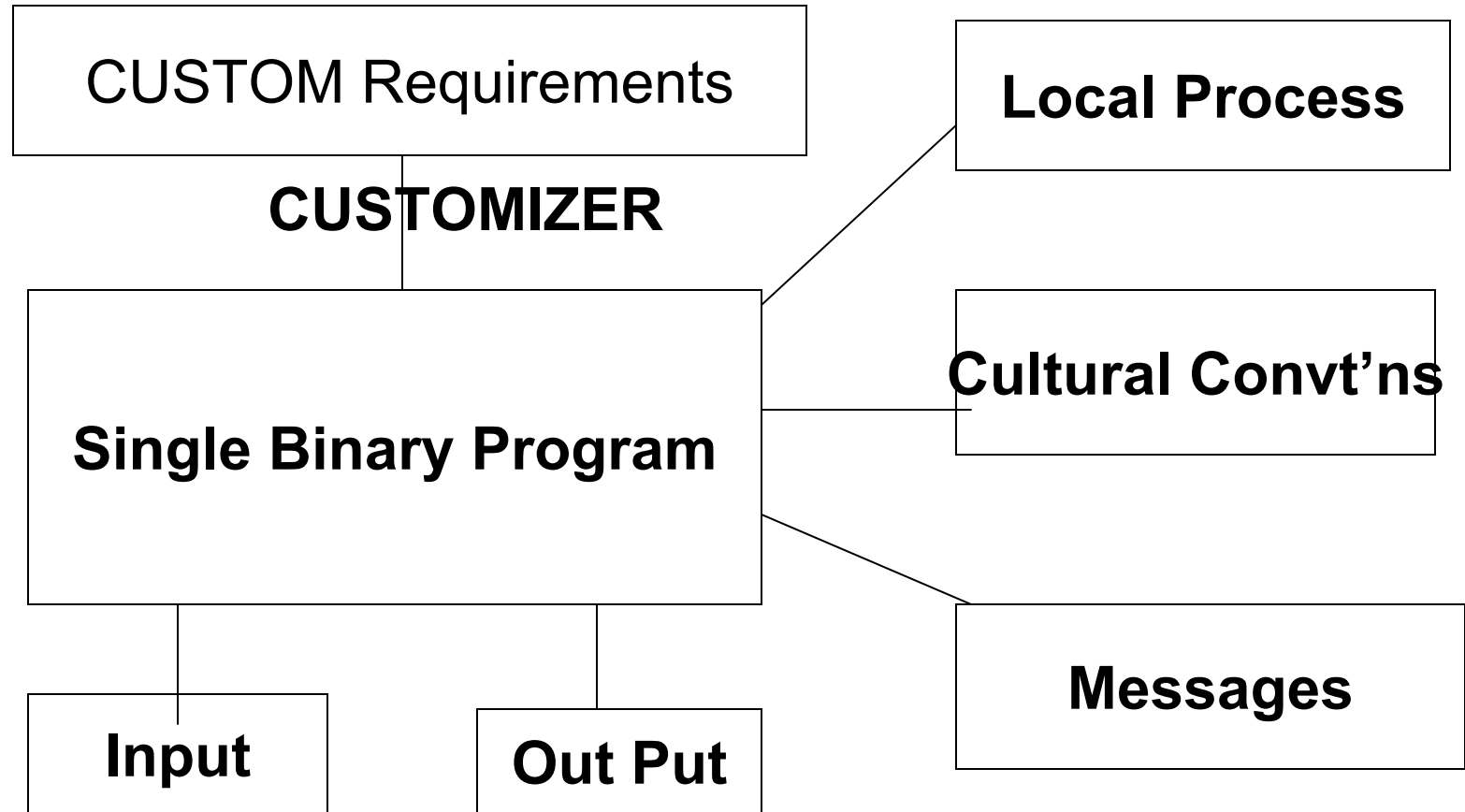
every pieces are costly and time consuming

Modification causes incompatibility and
Quality Problems

Maintenance Issue

**Maintenance should be done by local
including backward compatibility**

I18N/L10N model



All Cultural Dependents should be outside main program

“Single Binary” is a way to go

- Do NOT include any cultural dependent items within a program
- Do NOT include any character set dependency within a program
- All cultural dependencies should be at separated places from a program

Related Standards

- ISO/IEC 646
- ISO/IEC 2022
- ISO/IEC 2375
- ISO/IEC 6429
- ISO/IEC 8859
- ISO/IEC 10646
- ISO/IEC TR 15285
- ISO/IEC 10176
- ISO/IEC TR 11017
- ISO/IEC 14651
- ISO/IEC 14652
- ISO/IEC 15897
- Unicode

Related Organizations

- ISO
- IEC
- ITU-T
- ISO/IEC JTC1
- ISO/IEC JTC1 SC2
- ISO/IEC JTC1 SC22/WG20
- ISO/IEC JTC1 SC34
- ISO/IEC JTC1 SC35
- CEN
- ANSI
- Unicode Inc
- Open Standard Group
- W3C
- CICC

Part-2

ISO/IEC 10646 (Unicode)

Key Message

ISO/IEC 10646 (Unicode) is a tool
to make Single Binary Solution
available

Single Binary

Is NOT available at free of cost

Is available only when there is NO
character set dependency

Any local character requires Character
set dependency, if unconditionally coded

Multi-Source, Multi-Binary traditional localization

Character set dependent source code modification

Source code modification means multi-source,
multi-source means multi binary.

Single Source, Multi-Binary hope until mid 90th

Character set dependent compilation means multi-binary

Single Binary Modern localization

Write program one time, then, compile one time

Character sets have to have common behavior

No cultural dependent behavior allowed

Key elements of Coded Character Set

- Coding Scheme (container)
- Coded elements (characters on code)
- Behavior of the coded elements
- Coding order
- Naming of coded elements

Binary code is able to share (Single Binary) only when.....

- Coding scheme is the same
- Behavior of coded elements are the same
- The same elements with different behavior can not share the binary code

ISO/IEC 2022 approach does not meet for Single Binary requirement

- Locking Shift is difficult to process (traditional concern)
- Code developers have a free hands to select the behavior (multi-behavior processing needed)
- No way to predict new behavior

ISO/IEC 10646 approach

- Wide coding Space no locking shift
- Common Behavior

Compromise on coding is needed
This is very unpopular for many people

Common Behavior

- Unification
- Character and Glyph separation
- Normalization, Combining Sequence

Ambiguity Possibility Issue
Dual encoding issue

Ambiguity of sequence

There are three different characters

木 林 森

木 木 木

木木木？ 林木？ 木林？ 森？

木 NSNJ 木木

木 林

Frequently raised Objections

- Character naming
 - Unnecessary characters
 - Coding sequence
 - Printing Shape
-
- Compromise on coding, not on program is an design objective of ISO/IEC 10646

Is ISO/IEC 10646 perfect?

Are there other choices?

What is industry standard?

It is a choice,
It is not a truth

Answers for most of the objection are
“So What?”

Part-3 Input Method

A hint for much friendly Asian input methods

Key Messages

Do NOT standardize input method
in such a hurry.

Improvement for better user
friendliness should be continued

No need to standardize
a keyboard layout now

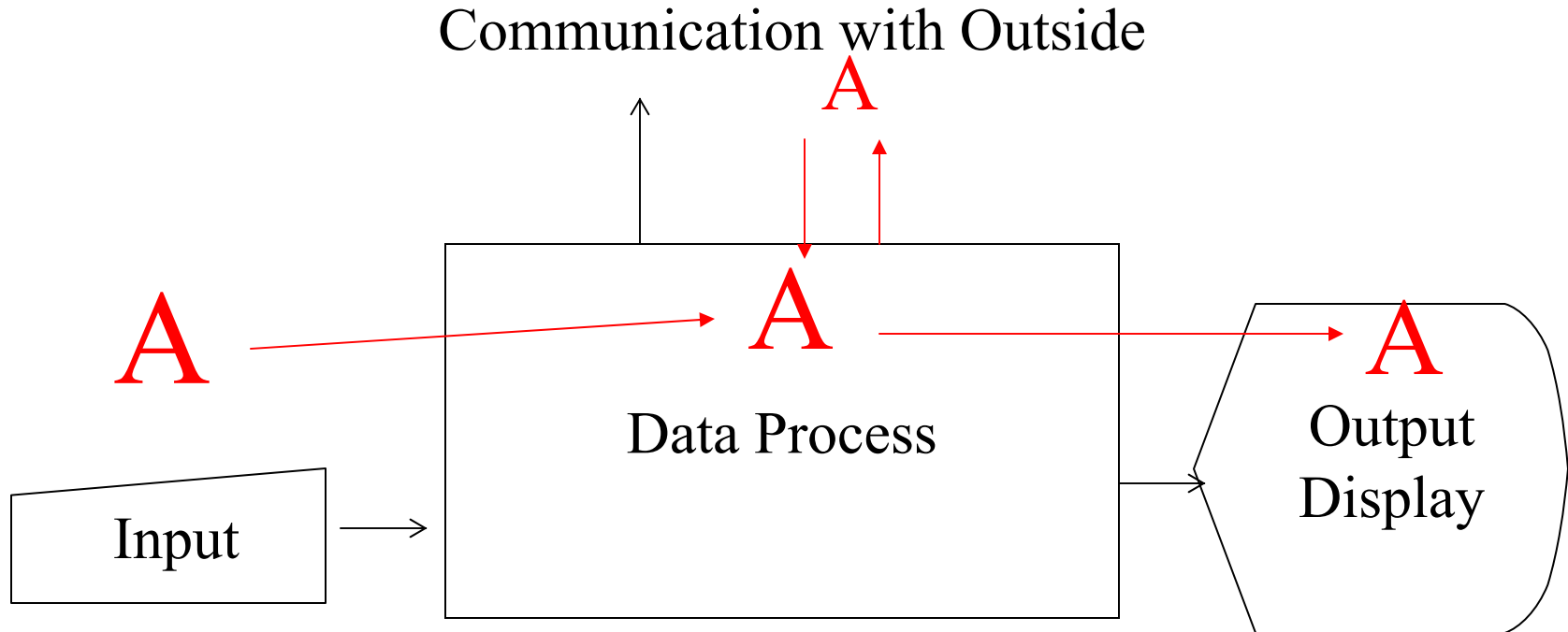
Functionality requirements for input
is “code stream generation”

Starting Point of
Ubiquitous Computing is Input

Japanese Input Method

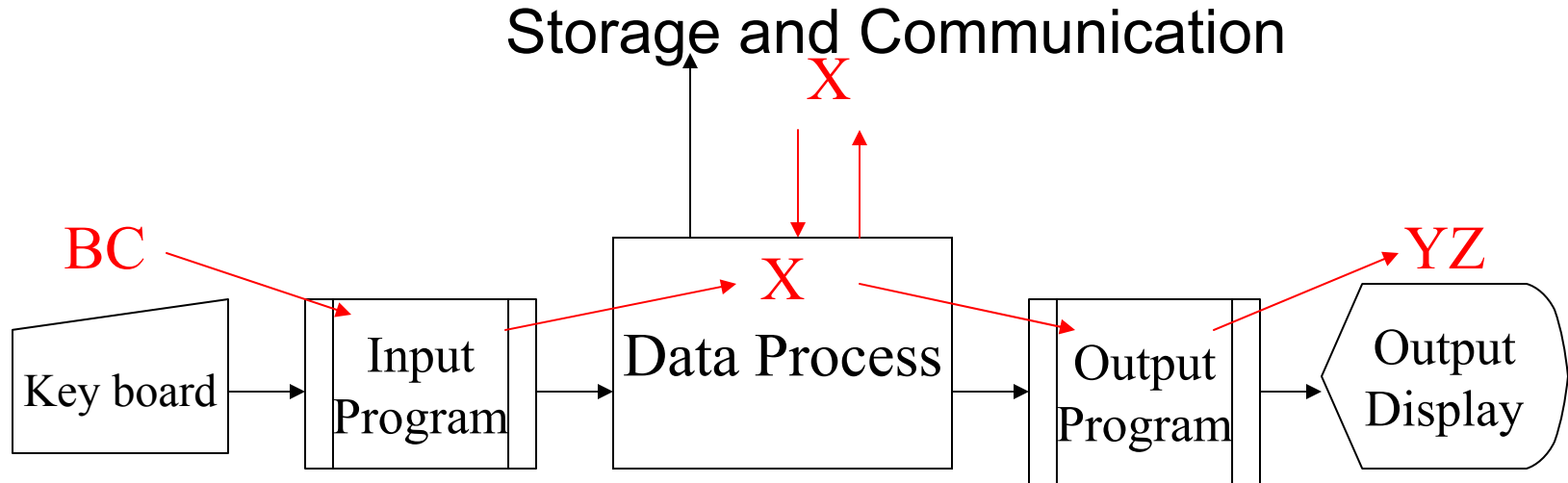
- Input 7000++ characters by using regular keyboard. How?
- Key stroke is not equal to code stream
- Multiple key strokes generate single code
- Small number of key stroke generates long string of codes

Old Style Input



- Same character every where
- One change at somewhere change on Data Process part
- Change in data process means incompatibility

Modern Input Method



- Input Program and Output Program isolate processes
- The Programs convert information to any form
- Thus, Data at data process are kept consistent
- While, Any input sequence or any output possible
- Continuous upgrade for better user interface possible

Hint

Is formatting characters friendly
with users?

Conclusions (1)

- For cost effective localization, a Single binary is the way to go
- To make the single binary, character code should make compromise to fit to program
- Fonts and Input Methods should absorb the compromised results for user friendliness
- International standard is needed to disclose the local requirements

Conclusions (2)

- Localization is costly business
- Localization is the long term commitment
- Avoid naïve localization and naïve outside consulting and assistant
- Encourage (independent) local developments
- Do not do over standardization
- National strategy for localization as a part of IT-Policy is the “MUST”
- Utilize the outside assistance as a part of the element of rode-map of the strategy

What is CICC?

see different slides

Thanks