



# **Open Source Project Universal Locales for Linux**

**- Locale data in XML and Transformer**

**Kentaro Noji**

**Globalization Center of Competency, Yamato**

**Yamato Software Laboratory, IBM Japan Ltd.**

# Why Universal Locales for Linux

- ✦ Linux is popular OS now...
- ✦ Linux needs to improve globalization...
- ✦ **Solution**
  - ✦ Li18nux:2000 Globalization Spec.
  - ✦ Deployment of Unicode technology for Linux

# Objectives

- ★ **Deployment of Unicode Technology**
  - ✱ Usage of Unicode resources
  - ✱ Apply Unicode-based locale data in existing resources to Linux
- ★ **Platform independent framework**
  - ✱ Locale data consistency
  - ✱ Usage of XML for locale data definition
- ★ **Open source license**
  - ✱ For all Unicode user on Linux

# Unicode Technical Resources

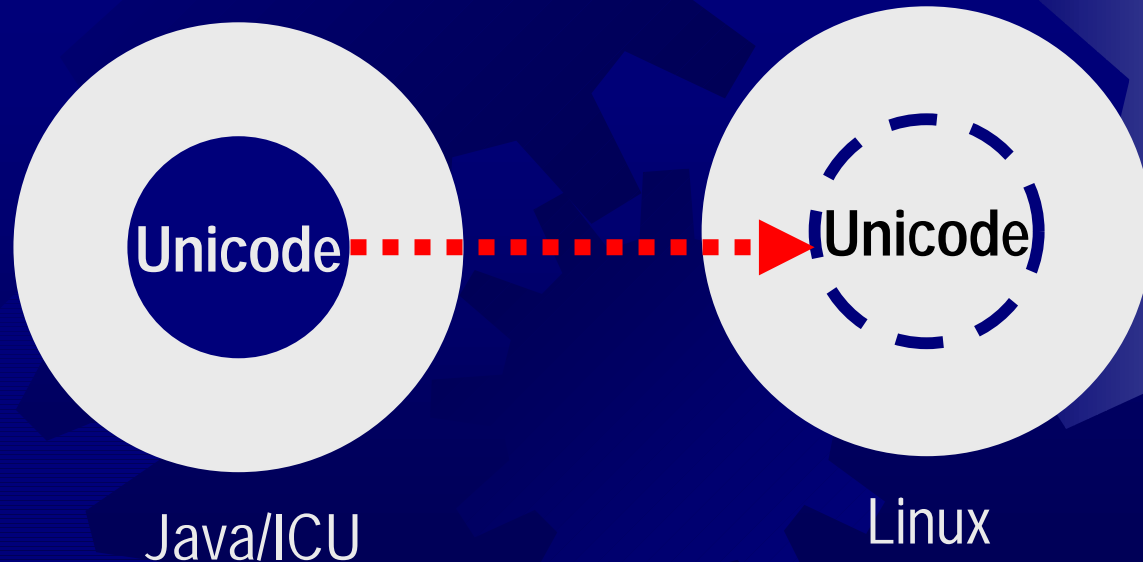
## ★ Deployment of Unicode Std. to Linux

- ★ Character Set : Unicode V3.0
- ★ Encoding : UTF-8
- ★ Collation
  - ★ Unicode Collation Algorithm – UTS#10
- ★ Character Classification
  - ★ Unicode character data base

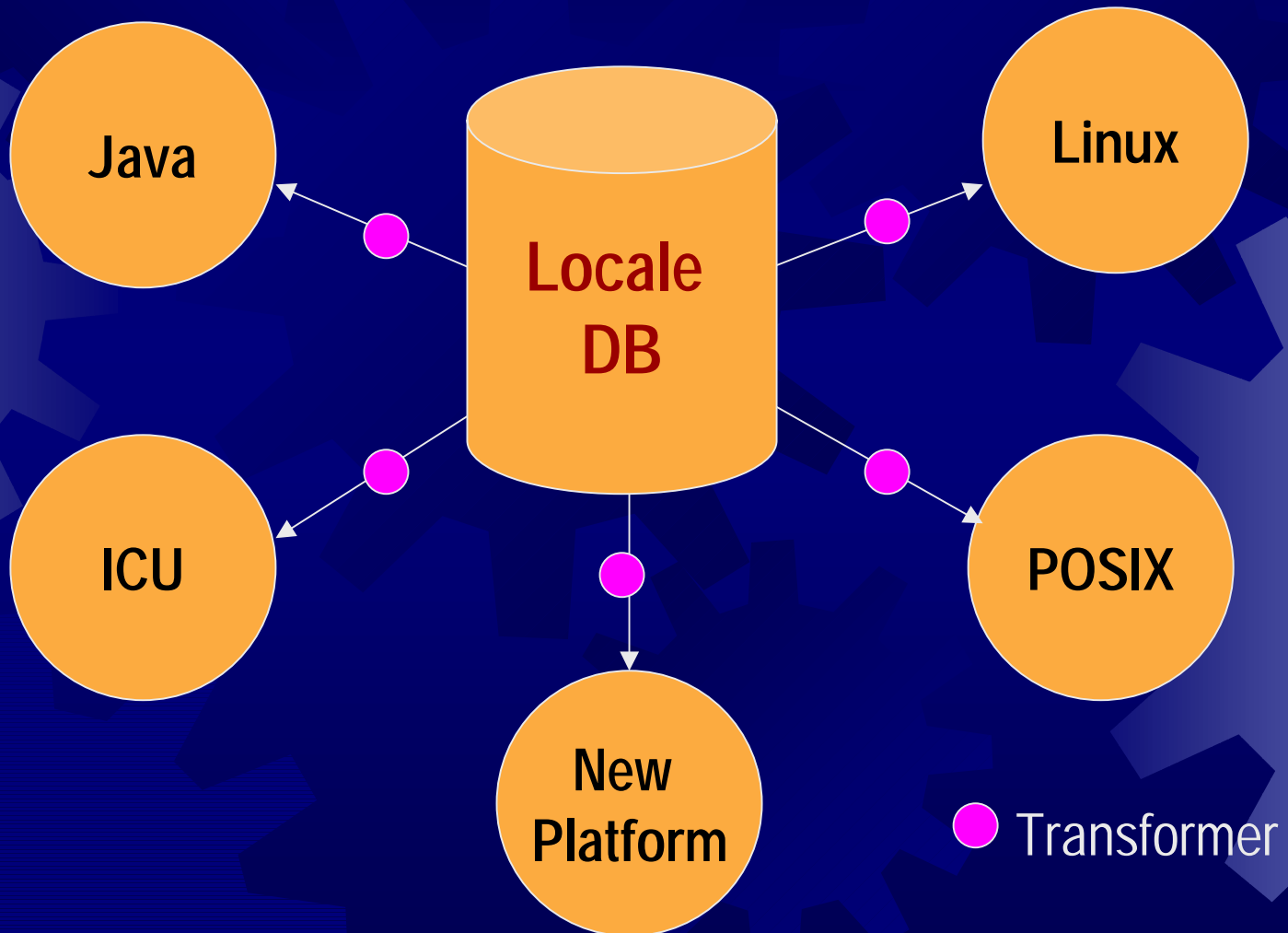
# Unicode Locale Data

- ★ ICU architecture provides Unicode-based locale for Java and ICU for C/C++
- ✿ Co-work with ICU development

ICU: International Components for Unicode



# Platform independent



# XML locale data and Transformer

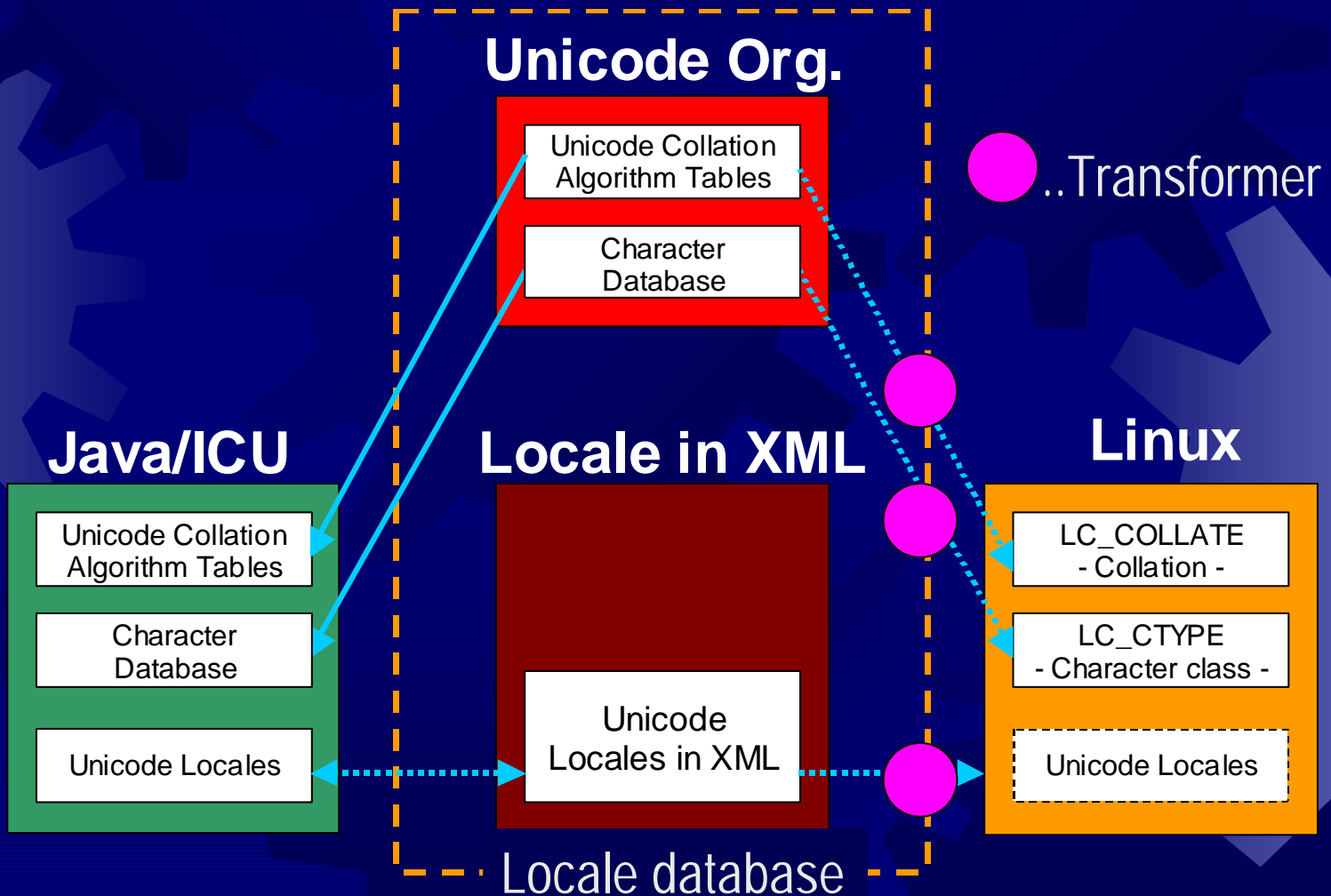
- ✦ **XML Locale data**

- ✦ For platform independent

- ✦ **Transformer**

- ✦ Converts the XML locale data to a platform unique locale data format
  - ✦ E.g from XML to Linux locale source

# Universal Locales project





## Example – Currency locale data in XML

```
<?xml version="1.0" encoding="UTF-8" ?>  
  
...  
<currencies>  
  <currency id="JPY" default="true">  
    <symbol> ¥ </symbol>  
    <name>JPY</name>  
    <decimal>.</decimal>  
    <pattern>α#,##0</pattern>  
  </currency>  
</currencies>  
  
...
```

# Locale Data in XML

- ★ Easy to support new platforms by adding a transformer
- ★ Easy to verify locale data
- ★ Easy to add new locale data
- ★ Easy to make locale data consistent among various platforms

# Transformer

## ★ Three transformations

- ★ XML locale data to POSIX style locale data except collation and character classification
- ★ Unicode Collation keys to POSIX style collation table
- ★ Unicode character database to POSIX style character classification table

# Example – Transformation from XML to Linux

## XML-based Locale

```
<currencies>
<currency id="JPY"
<symbol>¥ </symbol>
<name>JPY</name>
<decimal>.</decimal>
<pattern>¤#,##0</pa
</currency>
</currencies>
```

## Universal Locales for Linux

```
LC_MONETARY
int_curr_symbol
    "<U004A><U0050><U0059><U0020>"
currency_symbol
    "<UFFE5>"
mon_decimal_point
    "<U002E>"
mon_thousands_sep
    "<U002C>"
mon_grouping
    3
positive_sign
    ""
negative_sign
    "<U002D>"
int_frac_digits
    0
frac_digits
    0
p_cs_precedes
    1
p_sep_by_space
    0
n_cs_precedes
    1
n_sep_by_space
    0
p_sign_posn
    1
n_sign_posn
    1
END LC_MONETARY
```

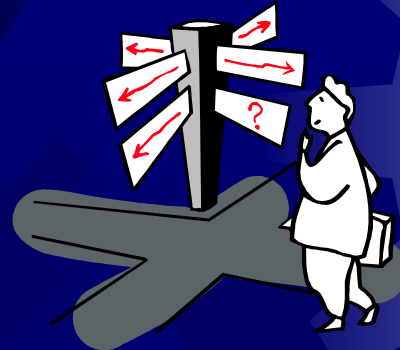
Transformer

# Conclusion: What we have done

- ★ Used XML to define locale data
  - ★ Platform independent
  - ★ Locale data consistent
- ★ Developed transformer
- ★ Made Unicode-based Linux Locales available
  - ★ 146 locales
  - ★ IBM developerWorks
  - ★ Li18nux
  - ★ ...

# What we will: 2001

- ★ **Add Indian Locales**
  - **Calling for contributors!**
- ★ **Collation enhancement**
  - **Support New UCA table**
  - **Add Tailoring**
- ★ **Add demonstration**
  - **Locale Walker**



# Demonstration

## ☀ Locale Walker

- ☀ Web Browser-based Locale data viewer
  - ☀ Month, Week, Date and Time format
  - ☀ Currency data and format
  - ☀ Numeric data and format
  - ☀ Collation
  - ☀ Character Classification



File Edit View Search Go Bookmarks Tasks Help Debug QA




Search



Bookmarks Home News Downloads Software Hardware Developers Help Search Shop

# Locale Walker

Universal locales for Linux

## Locale List

[ja\\_JP](#)

## About ja\_JP.UTF-8 locale

Locale	ja_JP.UTF-8
Encoding	UTF-8
Language	Japanese
Territory	Japan
Source	IBM Corporation
Revision	1.00

## Locale Categories

LC_TIME	Month Day Date and Time Format
LC_MONETARY	Monetary(currency) Format
LC_NUMERIC	Numearic Format
LC_MESSAGES	Yes and No message
LC_COLLATION	Collation demo
LC_CTYPE	Character classification demo

 Open yes not message data window

Mozilla












# Locale Walker

Universal locales for Linux

## Locale List

[ja\\_JP](#)

# Locale Walker

Universal locales for Linux

## Japanese, Japan

### Locale List

### Day data (LC\_TIME in ja\_JP.UTF-8 locale)

[ja\\_JP](#)

Day

	Name	Abbreviated Name
1st Day	日曜日	日
2nd Day	月曜日	月
3rd Day	火曜日	火
4th Day	水曜日	水
5th Day	木曜日	木
6th Day	金曜日	金
7th Day	土曜日	土

Open



File Edit View Search Go Bookmarks Tasks Help Debug QA



http://localhost/cgi-bin/localewalker?ja\_JP.UTF-8

Search



Bookmarks Home News Downloads Software Hardware Developers Help Search Shop

# Locale Walker

Universal locales for Linux

## Locale List

Japanese

Faeroese

Finnish

French

Galician

German

Greek

Greenlandic

Hebrew

Hindi

Hungarian

Icelandic

Indonesian

Irish

Italian

Japanese

Korean

Latvian

Lithuanian

Macedonian

Maltese

Norweigian

## About ja\_JP.UTF-8 locale

Locale	ja_JP.UTF-8
Encoding	UTF-8
Language	Japanese
Territory	Japan
Source	IBM Corporation
Revision	1.00

## Locale Categories

LC_TIME	Month Day Date and Time Format
LC_MONETARY	Monetary(currency) Format
LC_NUMERIC	Numearic Format
LC_MESSAGES	Yes and No message
LC_COLLATION	Collation demo
LC_CTYPE	Character classification demo

Open Month data window

Mozilla



[root@ymtk...]

★ Locale Wal...

[root@ymtk...]

The GIMP



File Edit View Search Go Bookmarks Tasks Help Debug QA




Search


[Bookmarks](#) [Home](#) [News](#) [Downloads](#) [Software](#) [Hardware](#) [Developers](#) [Help](#) [Search](#) [Shop](#)

# Locale Walker

Universal locales for Linux

## Locale List

[ko\\_KR](#)

## About ko\_KR.UTF-8 locale

Locale	ko_KR.UTF-8
Encoding	UTF-8
Language	Korean
Territory	Republic of Korea
Source	IBM Corporation
Revision	1.01

## Locale Categories

LC_TIME	Month Day Date and Time Format
LC_MONETARY	Monetary(currency) Format
LC_NUMERIC	Numearic Format
LC_MESSAGES	Yes and No message
LC_COLLATION	Collation demo
LC_CTYPE	Character classification demo

Mozilla








File Edit View Search Go Bookmarks Tasks Help Debug QA



http://localhost/cgi-bin/localewalker?ko\_KR.UTF-8

Search



Bookmarks Home News Downloads Software Hardware Developers Help Search Shop

# Locale Walker

Universal

## Locale List

Korean

[ko\\_KR](#)

# Locale Walker

Universal locales for Linux

## Korean, Republic of Korea

### Locale List Date and Time format (LC\_TIME in ko\_KR.UTF-8 locale)

Korean

[ko\\_KR](#)

#### Date and Time

Date Format 2001년 04월 23일 월요일

Time Format 04시 53분 23초 GMT

Date and Time Format 2001년 04월 23일 월요일 오전 04시53분23초 GMT

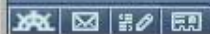
12 h Time Format 오전 04시 53분 23초 GMT

#### AM PM

AM 오전

PM 오후

Opera





# Locale Walker

Universal locales for Linux

## Korean, Republic of Korea

### Locale List

Korean

[ko\\_KR](#)

### Locale List

Korean

[ko\\_KR](#)

### Collation

(LC\_COLLATE in **ko\_KR.UTF-8** locale)

#### Input strings

#### Before Collation

#### After Collation

Submit

Reset

Sample A

Sample B

0 strings

0 strings

# Locale Walker

Universal locales for Linux

## Korean, Republic of Korea

**Locale List**      **Collation**  
(LC\_COLLATE in **ko\_KR.UTF-8** locale)

Korean

ko KR

Korean

ko KR

Input strings	Before Collation	After Collation
ou lésé péché vice-président 9999 OÛ haïe coop caennais lèse dû air@@@ colon bohème rânâ		
<input type="button" value="Submit"/> <input type="button" value="Reset"/>		
<input type="button" value="Sample A"/>		
<input type="button" value="Sample B"/>		
	0 strings	0 strings

# Locale Walker Universal locales for Linux

## Korean, Republic of Korea

**Locale List**      **Collation**  
(LC\_COLLATE in **ko\_KR.UTF-8** locale)

**Locale List**

Korean

[ko\\_KR](#)

Korean

[ko\\_KR](#)

**Input strings**

**Before Collation**

**After Collation**

ou	@@@@@
lésé	@@@air
péché	0000
vice-président	9999
9999	Aalborg
OÙ	aide
haïe	aïeul
coop	air
caennais	air@@@
lèse	Ålborg
dû	août
air@@@	bohème
côlon	Bohème
bohème	Bohémien
géné	C.A.F.
lamé	caennais
pêche	çàetlà
LÉS	Canon
viceversa	cañon
C.A.F.	casanier
cæsius	cæsius
resumé	co-op
Bohémien	CO-OP
co-op	colon
pécher	côlon

Submit    Reset

Sample A

Sample B

# URL

<http://oss.software.ibm.com/developerworks/opensource/locale/>



