

MigrateToICU

Migrate locale data to icu to remove the overlap in luni

Luni

Class	issue	Status	Comments
CurrencyTest	4 failures because of different locale data between ri and icu	Done	ICU4j 3.8 is CLDR 1.5 compatible whereas RI uses the old version
FormatterTest	1 failure because icu has different data in Japan locale	Done	Format pattern: %tb Argument: 0 expected:<1> but was:<1>, this is also caused by CLDR gap
GregorianCalendarTest	1 failure because icu has different data about the MinimalDaysInFirstWeek	Reported to ICU	test_getMinimalDaysInFirstWeek Failure expected:<4> but was:<1>, this is also caused by CLDR gap
LocaleTest	6 failures because of different locale data between ri and icu	Done	regression test for HARMONY-1146 failed because icu does not support the legacy countries. One difference about the variant has been reported as ICU ticket 6096
ScannerTest	2 failures in parsing float value in locale ar_AE	Ongoing	even more testcases fail on RI ...

Text

CollationElementIteratorTest	3 failures	Done	Fails on RI, correct the test	en
DecimalFormatSymbolsTest	1 failure for serialization compatibility	Ongoing	This test fails on RI too	
NumberFormatTest	1 failure for different data in ar_AE	Done	expected:<36-> but was:<->. I suppose the 36 could be represented as in this locale...google translator proved it. http://translate.google.com/translate_t?langpair=ar	
RuleBasedCollatorTest	2 failures	Done	Fails on RI, correct the test	
SimpleDateFormatTest	9 failures since ICU has different pattern against RI	Done	see below	

Differences in DateFormat/SimpleDateFormat

1. Icu supports extra pattern chars[1], refer to <http://www.icu-project.org/apiref/icu4j/com/ibm/icu/text/SimpleDateFormat.html>; most failures in `test_applyLocalizedPatternLjava_lang_String()` are caused by that.
2. For pattern "ZZZZ", ICU has different output, for example icu returns GMT-07:00 but RI returns -7000.
3. For pattern "z" and "zzzz", ICU has different output, for example icu returns "GMT+14:00, Line Islands Time" for "z, zzzz" but RI returns "LINT, Line Is. Time" in "Pacific/Kiritimati"; most failures in `test_timeZoneFormatting()` are caused by above 2.
4. `test_formatLjava_util_DateLjava_lang_StringBufferLjava_text_FieldPosition()` reported as ICU ticket 6089.
5. icu has different explanation on the pattern 'y'.

To be continued.

[1]

Symbol	Meaning	Presentation	Example
-----	-----	-----	-----
G	era designator	(Text)	AD
y†	year	(Number)	1996
Y*	year (week of year)	(Number)	1997
u*	extended year	(Number)	4601
M	month in year	(Text & Number)	July & 07
d	day in month	(Number)	10
h	hour in am/pm (1-12)	(Number)	12
H	hour in day (0-23)	(Number)	0
m	minute in hour	(Number)	30
s	second in minute	(Number)	55
S	fractional second	(Number)	978
E	day of week	(Text)	Tuesday
e*	day of week (local 1-7)	(Number)	2
D	day in year	(Number)	189
F	day of week in month	(Number)	2 (2nd Wed in July)
w	week in year	(Number)	27
W	week in month	(Number)	2
a	am/pm marker	(Text)	PM
k	hour in day (1-24)	(Number)	24
K	hour in am/pm (0-11)	(Number)	0
z	time zone	(Text)	Pacific Standard Time
Z	time zone (RFC 822)	(Number)	-0800
v	time zone (generic)	(Text)	Pacific Time
g*	Julian day	(Number)	2451334
A*	milliseconds in day	(Number)	69540000
Q*	quarter in year	(Text & Number)	Q1 & 01
c*	stand alone day of week	(Text & Number)	Tuesday & 2
L*	stand alone month	(Text & Number)	July & 07
q*	stand alone quarter	(Text & Number)	Q1 & 01
'	escape for text	(Delimiter)	'Date='
''	single quote	(Literal)	'o'clock'

* These items are not supported by Java's SimpleDateFormat.

† ICU interprets a single 'y' differently than Java.

Related JIRAs

ID	Description	Status
<ac:structured-macro ac:name="unmigrated-wiki-markup" ac:schema-version="1" ac:macro-id="fe1ed0a2-0696-4e2b-8cfd-86d2a99dc633"><ac:plain-text-body><![CDATA[	5085	[classlib][swing] [InternationalFormatter]RTest.testIncrementDecrement failure
<ac:structured-macro ac:name="unmigrated-wiki-markup" ac:schema-version="1" ac:macro-id="80fbbce2-3483-4812-88a9-a3a7d0311c21"><ac:plain-text-body><![CDATA[	5129	[classlib][icu] text formatting degradation after at 592434
<ac:structured-macro ac:name="unmigrated-wiki-markup" ac:schema-version="1" ac:macro-id="589c0600-71c0-442c-8960-0f608237a9b9"><ac:plain-text-body><![CDATA[	5122	[classlib][icu] significant performance degradation atfter ICU related commits
<ac:structured-macro ac:name="unmigrated-wiki-markup" ac:schema-version="1" ac:macro-id="1f500362-4d03-4320-9004-82364c07fc7c"><ac:plain-text-body><![CDATA[	5061	[classlib][luni][text]migrate locale data to ICU4j's

<ac:structured-macro ac:name="unmigrated-wiki-markup" ac:schema-version="1" ac:macro-id="9b704752-f23f-4cd3-acc4-f984699108be"><ac:plain-text-body><![CDATA[	5100	[classlib]Dacapo.jython regression after migrating the locale data	Closed	]]></ac:plain-text-body></ac:structured-macro>
<ac:structured-macro ac:name="unmigrated-wiki-markup" ac:schema-version="1" ac:macro-id="c7206a3e-cefd-4921-a3f3-5c4833b533f"><ac:plain-text-body><![CDATA[	5406	[classlib]CLDR data in harmony is different from that in RI	Open	]]></ac:plain-text-body></ac:structured-macro>
<ac:structured-macro ac:name="unmigrated-wiki-markup" ac:schema-version="1" ac:macro-id="feaf73d8-fdbd-4948-befe-c45541b59ada"><ac:plain-text-body><![CDATA[	5466	[classlib][util] display language is reset to the value of constructor after Locale is set as default	Open	]]></ac:plain-text-body></ac:structured-macro>
<ac:structured-macro ac:name="unmigrated-wiki-markup" ac:schema-version="1" ac:macro-id="4b9a71f9-6157-4eb5-b749-3f5b79f9bf19"><ac:plain-text-body><![CDATA[	5467	[classlib][util] incompatible country names for certain languages	Open	]]></ac:plain-text-body></ac:structured-macro>
<ac:structured-macro ac:name="unmigrated-wiki-markup" ac:schema-version="1" ac:macro-id="9bf75ab1-255d-49d5-b1d1-54eff6274f36"><ac:plain-text-body><![CDATA[	5448	[classlib][luni] [FormatterTest] will fail beacuse of different locale data between ri and icu	In Progress	]]></ac:plain-text-body></ac:structured-macro>
<ac:structured-macro ac:name="unmigrated-wiki-markup" ac:schema-version="1" ac:macro-id="d49e4213-31f9-4681-861a-ce7974d50629"><ac:plain-text-body><![CDATA[	5442	[classlib][util] some testcases of org.apache.harmony.luni.tests.java.util.LocaleTest fails on RI	Closed	]]></ac:plain-text-body></ac:structured-macro>
<ac:structured-macro ac:name="unmigrated-wiki-markup" ac:schema-version="1" ac:macro-id="e5ea99f0-e56a-4bdc-acee-67d24bd91a52"><ac:plain-text-body><![CDATA[	5439	[classlib][luni] test_getMinimalDaysInFirstWeek in org.apache.harmony.luni.tests.java.util.GregorianCalendarTest fails on RI	Closed	]]></ac:plain-text-body></ac:structured-macro>
<ac:structured-macro ac:name="unmigrated-wiki-markup" ac:schema-version="1" ac:macro-id="25121790-eef9-427b-975b-db0cd7ff2a6a"><ac:plain-text-body><![CDATA[	5407	[classlib][luni]4 testcases in java.util.CurrencyTest fails on RI	Closed	]]></ac:plain-text-body></ac:structured-macro>
<ac:structured-macro ac:name="unmigrated-wiki-markup" ac:schema-version="1" ac:macro-id="5cc784db-1a1c-4ee3-8f68-51da26fd02a7"><ac:plain-text-body><![CDATA[	5467	[classlib][util] incompatible country names for certain languages	Open	]]></ac:plain-text-body></ac:structured-macro>
<ac:structured-macro ac:name="unmigrated-wiki-markup" ac:schema-version="1" ac:macro-id="b5014a9d-f59f-4400-b8f2-ef11d6f870c6"><ac:plain-text-body><![CDATA[	5459	[classlib][util] default [TimeZone] ignores raw offset after switch to ICU	Fixed	]]></ac:plain-text-body></ac:structured-macro>
<ac:structured-macro ac:name="unmigrated-wiki-markup" ac:schema-version="1" ac:macro-id="c70edb28-5070-46ca-88fa-ef6ce065fc8f"><ac:plain-text-body><![CDATA[	5461	[classlib][util] wrong daylight savings are reported for "MST" and "EST" timezones after locale data unification	Open	]]></ac:plain-text-body></ac:structured-macro>
<ac:structured-macro ac:name="unmigrated-wiki-markup" ac:schema-version="1" ac:macro-id="8b6415d0-8414-4dff-b13d-80ac509462b6"><ac:plain-text-body><![CDATA[	5465	[classlib][util] Locale(String, String, String) capitalize third parameter (variant)	Open	]]></ac:plain-text-body></ac:structured-macro>
<ac:structured-macro ac:name="unmigrated-wiki-markup" ac:schema-version="1" ac:macro-id="04d51fa9-10f5-40eb-8780-b321f67b52e6"><ac:plain-text-body><![CDATA[	5468	[classlib][util] java.util.Date.toString outputs real time zone name on RI, but prints GMT<+/-offset> on Harmony	In progress	]]></ac:plain-text-body></ac:structured-macro>
<ac:structured-macro ac:name="unmigrated-wiki-markup" ac:schema-version="1" ac:macro-id="d41b6d34-66d7-44f0-8c4e-34bac07d9636"><ac:plain-text-body><![CDATA[	5469	[classlib][util] [DateFormat].SHORT has different meaning on Harmony and RI	In progress	]]></ac:plain-text-body></ac:structured-macro>

ICU trac

6035 6036
6037

6043 6044
6045

6089 6095
6096

6174

In queue to report:
Harmony-5436 Harmony-5407

Others

593469 fix the regression that we have different locale data for Europe/London, along with some fixes for other difference against RI 593024 Fix a bug that [SimpleDateFormat](#).parse does not reload the latest timezone 592434 Apply patch Harmony-5061 which removes the duplicate locale data

reverted at 596044

patch recommitted at r612718

JIRA to record non-bug differences: