State of CJK issues of LibreOffice, 2018 edition
Shinji Enoki
shinji.enoki@gmail.com
TIRANA | 27 Sept. 2018
Agenda

- About the CJK features & issues
- CJK issue status
- Next challenge for the CJK issues
Who am I?

• Shinji Enoki (榎真治)
• A member of LibreOffice Japanese Team (2011-)
• A member of The Document Foundation (2014-)
• Activity: organizing events, grows community, sometimes QA
• Other community:
  - Chairperson of Japan UNIX Society
  - A member of Kyoto OpenData Community
  - A staff of KANSAI OPEN SOURCE
Do you know the word CJK bug?
What are CJK issues / bugs

- CJK is an abbreviation for Chinese-Japanese-Korean
- LibreOffice has many language-specific features and issues, CJK issue is one of them.
- CJK issues happen when used in CJK's environment
  - A function of CJK is broken
  - A bug occurs only with the CJK environment
Next Decade Manifesto

For the past ten years, the OpenOffice.org community has developed, supported and promoted the world's leading open-source office productivity suite. We have attracted the support of tens of thousands of individuals and corporate bodies during this period.

We now call on all our supporters to follow us into the next phase of our development, as we become an independent Foundation.

Our Values

WE COMMIT OURSELVES:

To eliminate the digital divide in society by giving everyone access to office productivity tools free of charge to enable them to participate as full citizens in the 21st century

To support the preservation of mother tongues by encouraging people to translate, document, support, and promote our office productivity tools in their mother tongue

To allow users of office productivity software to retain the intellectual property in the documents they create by use of open document formats and open standards

To an open and transparent peer-reviewed software development process where technical excellence is valued

WE REJECT:

The ownership of office productivity tools by monopoly suppliers which imposes a de-facto tax on global electronic free speech and penalises the economically disadvantaged

The creeping domination of computer desktops by a single language, forcing all people to learn a foreign language before they can express themselves electronically

The ownership of file formats by proprietary software companies - documents belong to their creators, not software vendors

A closed software development process where errors can lie hidden and poor quality is accepted
“To support the preservation of mother tongue by encouraging people to translate, document, support, and promote our office productivity tools in their mother tongue”

- I think these “Our Values” are very good
Why do CJK people need to work CJK bugs?

- Many LibreOffice developers are European people
- It is hard to notice / understand CJK bug unless it is from people of CJK
- Then, It is very important for CJK users to report / reproduce CJK bugs
• CJK bugs are tracked META issues
  - Bug 83066(CJK) - [META] issues
Bug 83066 (CJK) - [META] CJK (Chinese, Japanese, Korean, and Vietnamese) language issues

**Status:** NEW

**Alias:** CJK

**Product:** LibreOffice

**Component:** LibreOffice (show other bugs)

**Version:** unspecified

**Hardware:** Other All

**Importance:** medium normal

**Assignee:** Not Assigned

**URL:**

**Whiteboard:**

**Keywords:**

**Depends:** 35208 36709 39393 43671 50349 50901 55954 63011 65266 66791 68803 71329 72042 75790 77803 81272 81484 82115 82116 82293

**CC List:** 15 users (show)

**See Also:** Font-Rendering Vertical-Text RTL-CTL

**Crash report or crash signature:**
Check [META] CJK issues

<table>
<thead>
<tr>
<th></th>
<th>2017/10</th>
<th>2018/9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depends on:(All)</td>
<td>135</td>
<td>168</td>
</tr>
<tr>
<td>Open</td>
<td>59</td>
<td>73</td>
</tr>
<tr>
<td>Close</td>
<td>76</td>
<td>95</td>
</tr>
</tbody>
</table>

- Some cases are not added to CJK META issues due to CJK related bugs
• Mark Hung is fixing some CJK issues
• Thanks Mark!
CJK features

• In common with CJK
  1. using Kanji (Chinese characters)
  2. vertical writing
  3. phonetic guides (ruby)

• Chinese characters are also different in Simplified Chinese (Mainland China), Traditional Chinese (Taiwan, Hong Kong) and Kanji (Japanese)

• Even with common functions, usage methods are not always the same
  - Vertical writing
Vertical writing

- Example: Writer page style
  - [Page] tab
Vertical writing(2)
Features specific to Japanese

- Line breaking rules: "禁止処理" Kinsokushori
- [Paragraph style]-[Japanese appearance] tab

改行時の処理

- 行頭および行末の禁止文字のリストで禁止処理を行う(A)
- 句読点のぷら下げを行う(B)
- 日本語、英文字、および混在文章で文字間隔を調整(C)
Kanji variant is a Kanji character with the same meaning but by different shape.
Ideographic Variation Sequence/Selector

In Japan, using a mechanism called IVS to make it easier to deal with a Kanji variant.

IVS is combine the characters, as one of the character.
UCSコードは、細かな字形の差異を区別せず、複数の字形を一つのコードに統合している。

UCSコード

<table>
<thead>
<tr>
<th>字形</th>
<th>VS</th>
</tr>
</thead>
<tbody>
<tr>
<td>845A</td>
<td>E0100</td>
</tr>
<tr>
<td>845B</td>
<td>E0101</td>
</tr>
<tr>
<td>8FBA</td>
<td>E0102</td>
</tr>
</tbody>
</table>

UCSコードと字形を指定するVS（字形選択子）を並べて記述することで、字形を指定することができる。
※VSの値は実際の規格と異なります。
Deleting the character behind two same IVS characters with Backspace will result in another character

Demo
IVS using case bug#113481 (2)

• I explained this bug at LibreOffice conference 2017
• Mark Hung created a patch
• 6.1 : FIXED
Bug#35301: Allow vertical ruby text in the Asian phonetic guide.
Bug#35301 (2)

- LibreOffice 6.1 can become on the right side with ruby in horizontal writing
- Mark Hung created a patch
Bug 115258 - FILESAVE: text with Asian phonetic guide (ruby) disappear after save to docx and reopened

- Text with Asian phonetic guide (ruby) disappears after saving as “docx” and then reopening
- Mark Hung created a patch
- 6.1 : FIXED
Multilingual issues that matter even at CJK

- Multilingual problem that becomes a problem with CJK
- In some time zone Date is shifted: (Already fixed)
- Bug 59850 - EDITING: typed date decreased one day
- Bug 63230 - Calc: Date decreased by 1 depending on Timezone
Bug 118849 - Impress hang when input multibyte character.

- Only Linux
- 6.0.5 reproduced, 6.1 not reproduced
- demo
Meta issue for each CJK language

- [META] Japanese language-specific CJK issues
  - https://bugs.documentfoundation.org/show_bug.cgi?id=113195
- [META] Traditional Chinese (zh_TW, zh_HK)
  - https://bugs.documentfoundation.org/show_bug.cgi?id=113193
- [META] Simplified Chinese (zh_CN)
  - https://bugs.documentfoundation.org/show_bug.cgi?id=113194
- But, it is not used now
Checking is difficult

- It is difficult to understand the meaning
- Test case, sample file is not enough
- I do not understand when I read the comment to the end
Next task as I think

• 1st: Re-make or arrange the summary and test cases in Bugzilla
  - Against high risk and high probability of regression
• 2nd: CJK Meta issue and list maintenance
• 3rd: Finding critical bugs that users don't immediately notice
If you want to contribute to reporting or resolving CJK bugs, please join this channel:

https://web.telegram.org/#/im?p=s1119055114_15674911573967734518
Conclusion

- Sharing the Unique functions and issues about CJK with a lot of members
- I’d like to work to solve issues with our Japanese members
- wanting to progress organization of Bugzilla such as CJK issue
Thank you!