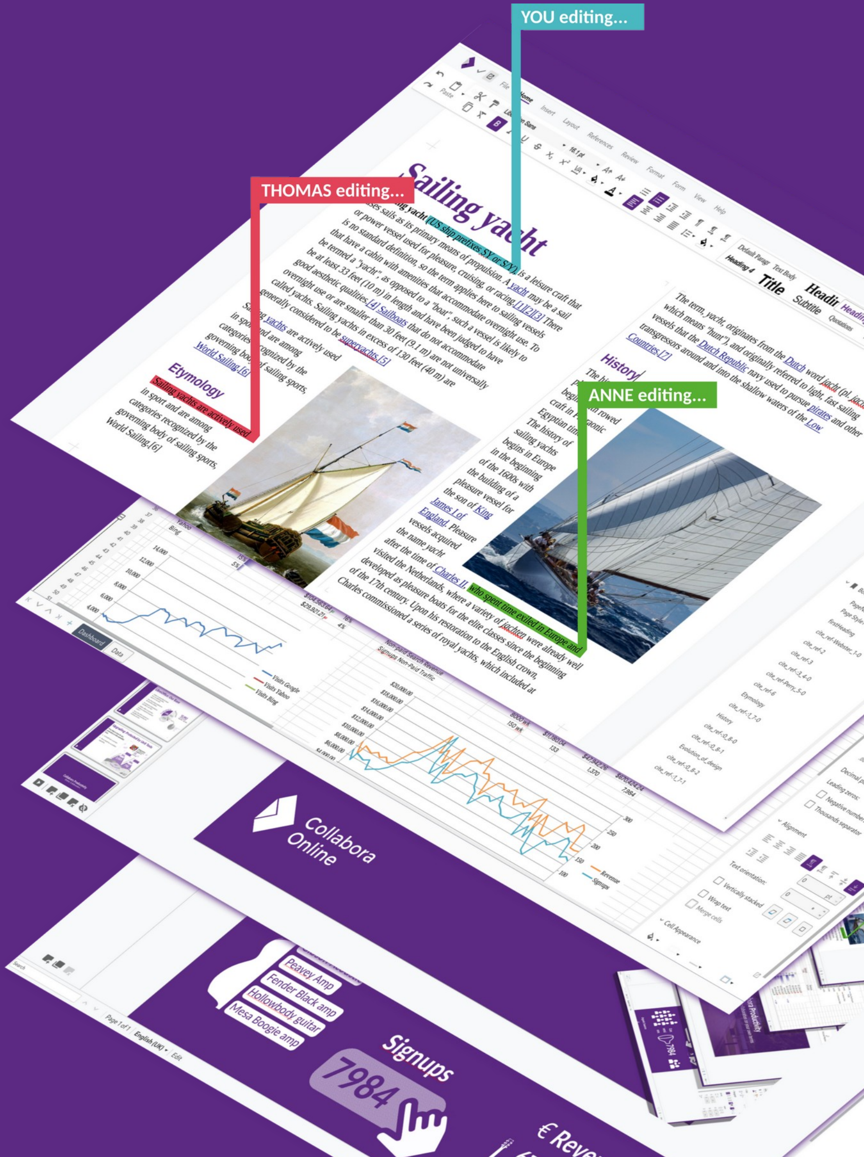
Tips for debugging your text documents

Miklos Vajna
Software Engineer
vmiklos@collabora.com



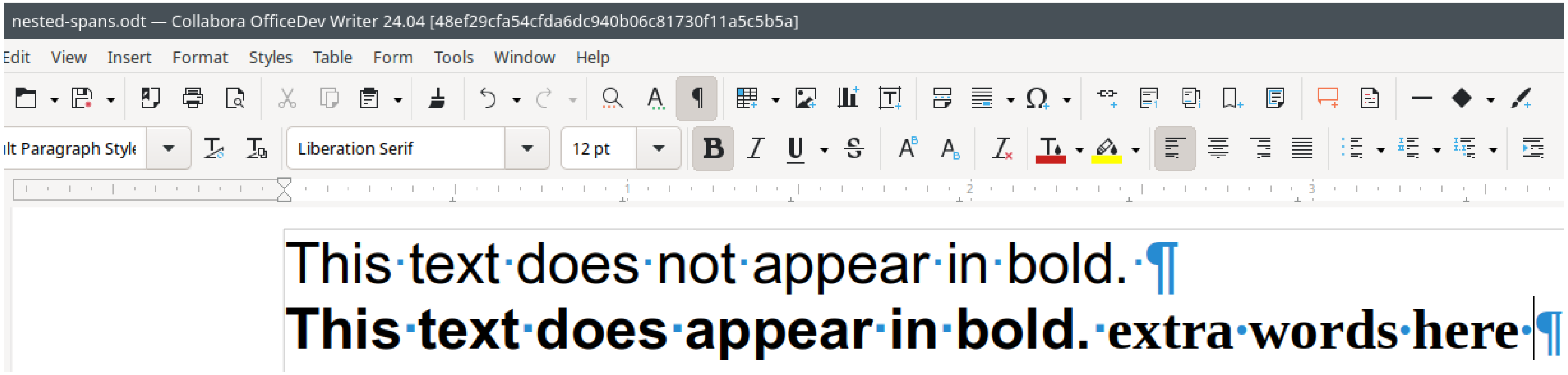
Collabora Online

LibreOffice Technology

Technical Day

COOL days

Document model level

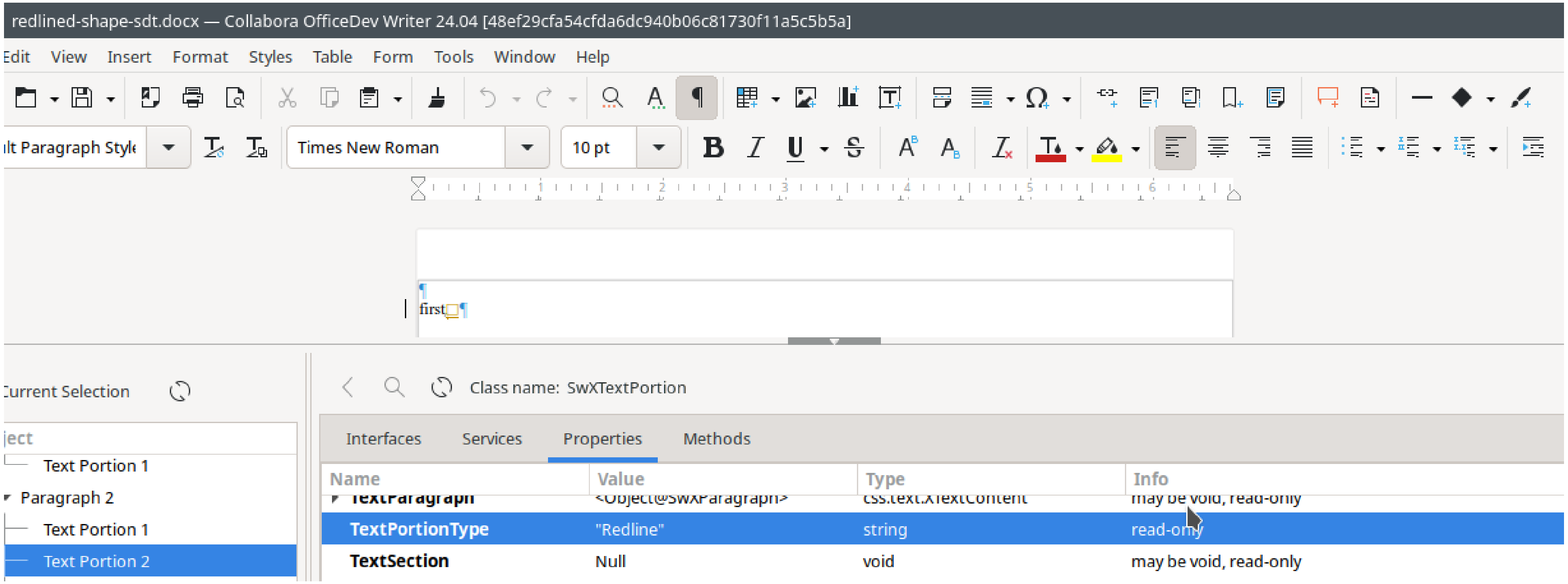

nested-spans.odt — Collabora OfficeDev Writer 24.04 [48ef29cfa54cfda6dc940b06c81730f11a5c5b5a]
Edit View Insert Format Styles Table Form Tools Window Help
lt Paragraph Style
Liberation Serif
12 pt

This·text·does·not·appear·in·bold.·¶
This·text·does·appear·in·bold.·**extra·words·here**·¶

- Can check what’s inside the C++ structures in the memory
- Set SW_DEBUG=1, then Shift-F12 produces a nodes.xml

Technical Day COOL days

UNO API level


redlined-shape-sdt.docx — Collabora OfficeDev Writer 24.04 [48ef29cfa54cfda6dc940b06c81730f11a5c5b5a]

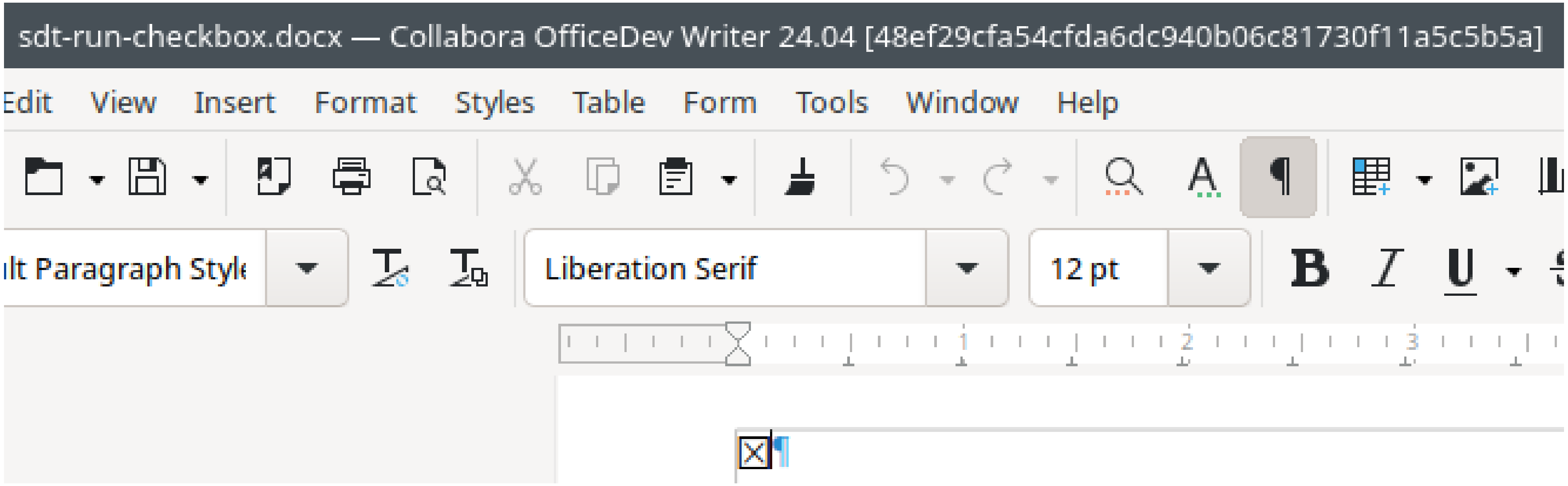
Edit View Insert Format Styles Table Form Tools Window Help
lt Paragraph Style
Times New Roman
10 pt
first
Current Selection
ject
Text Portion 1
Paragraph 2
Text Portion 1
Text Portion 2
Class name: SwXTextPortion
Interfaces Services Properties Methods

Name	Value	Type	Info
TextParagraph	<Object@SwXParagraph>	css.text.XTextContent	may be void, read-only
TextPortionType	"Redline"	string	read-only
TextSection	Null	void	may be void, read-only

- UNO API is visible to scripting, can write e.g. a (Python) macro
- Tools → Development tools in Collabora Office

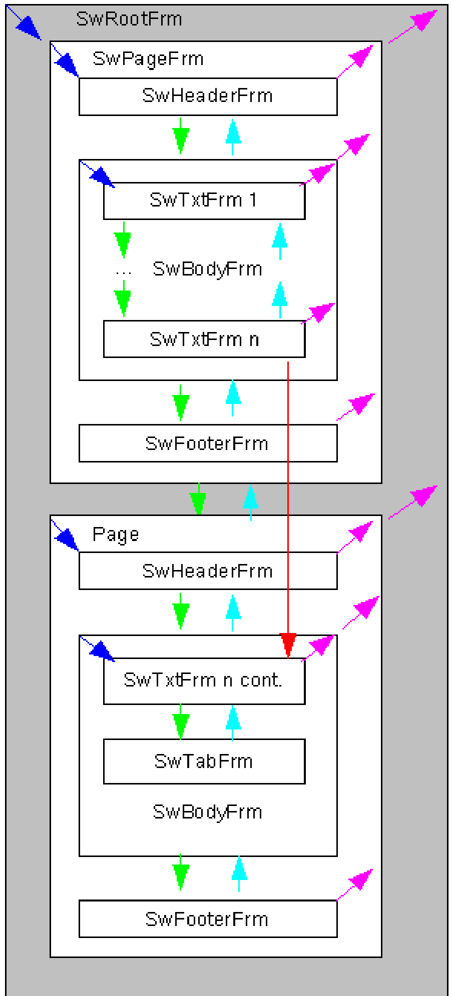
Collabora Online

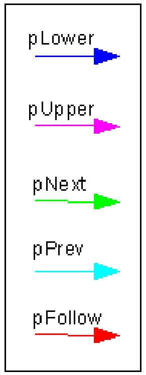
DOCX import level


sdt-run-checkbox.docx — Collabora OfficeDev Writer 24.04 [48ef29cfa54cfda6dc940b06c81730f11a5c5b5a]
Edit View Insert Format Styles Table Form Tools Window Help
lt Paragraph Style
Liberation Serif
12 pt

- DOCX import works by producing tokens, SW_DEBUG_WRITERFILTER=1 to see them in /tmp as a dump
- Based on that, it’s possible to decide if the bug is in the tokenizer or in the mapper (to UNO API calls)

Layout level

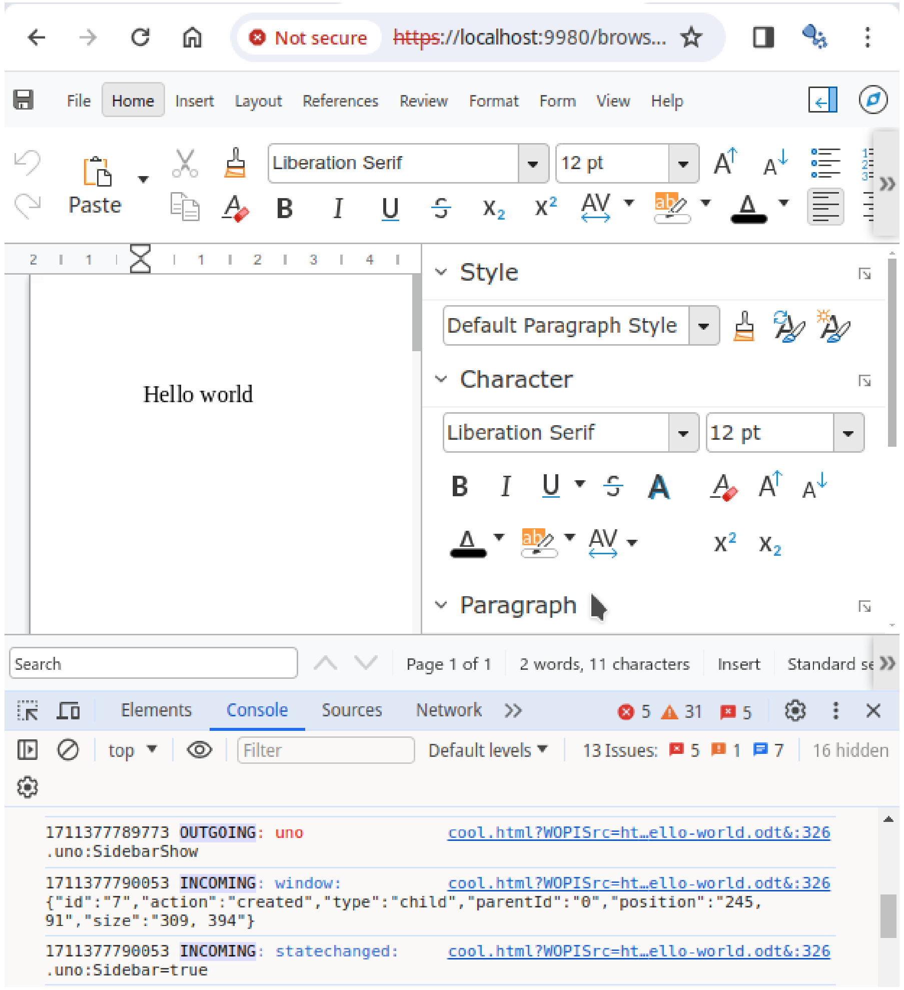




- Can check what's inside the frames in the memory
- Set SW_DEBUG=1, then F12 produces a layout.xml

Technical Day

UI / Online level



```
1711377789773 OUTGOING: uno                    cool.html?WOPISrc=ht…ello-world.odt&:326
.uno:SidebarShow
1711377790053 INCOMING: window:                cool.html?WOPISrc=ht…ello-world.odt&:326
{"id":"7","action":"created","type":"child","parentId":"0","position":"245,
91","size":"309, 394"}
1711377790053 INCOMING: statechanged:          cool.html?WOPISrc=ht…ello-world.odt&:326
.uno:Sidebar=true
```

- Can check outgoing protocol messages
- Can verify incoming responses

Collabora Online

Technical Day

COOL days

Summary

- Debugging text documents starts with finding which layer has the problematic behaviour: need to go wide
- Once that area is known, we can go deep in that layer
- Minimal, public reproducer documents are great
- Always solve the root cause, it pays off in the long run

Collabora Online