Re-structuring a giant, ancient code-base

or: Making LibreOffice work well everywhere

Michael Meeks

General Manager at Collabora Productivity

michael.meeks@collabora.com
Skype - mmeeks,
G+ - mejmeeks@gmail.com

"Stand at the crossroads and look; ask for the ancient paths, ask where the good way is, and walk in it, and you will find rest for your souls..." - Jeremiah 6:16

Re-structuring an overview

- Re-structuring for new platforms
- Making code readable
- Parallelising / threading: OCL / Threading
- Drawing & Widgets
  - primitives, gtk3/cairo, OpenGL, harfbuzz
- Security work: coverity, etc.
- Unit tests: protecting the backside ...
- Online
  - memory sharing
  - dialog tunnelling
- LibreOffice 6.0 bits
- Get involved

Why re-structure at all ?

- A 30+ year old code-base
  - Object Orientation mania
    - The silver-bullet
  - OO toolkit – less interesting
  - Demo / Office apps → very interesting.
- What languages should we use ?
- What platforms should we target ?
- *“Technology – the only industry that is more fashion driven than women’s fashion”*

Language choices:

Tip#1 - don’t buy a silver bullet:

"*there is no single development, in either technology or management technique, which by itself promises even one order of magnitude improvement within a decade in productivity, in reliability, in simplicity.*" Fred Brooks

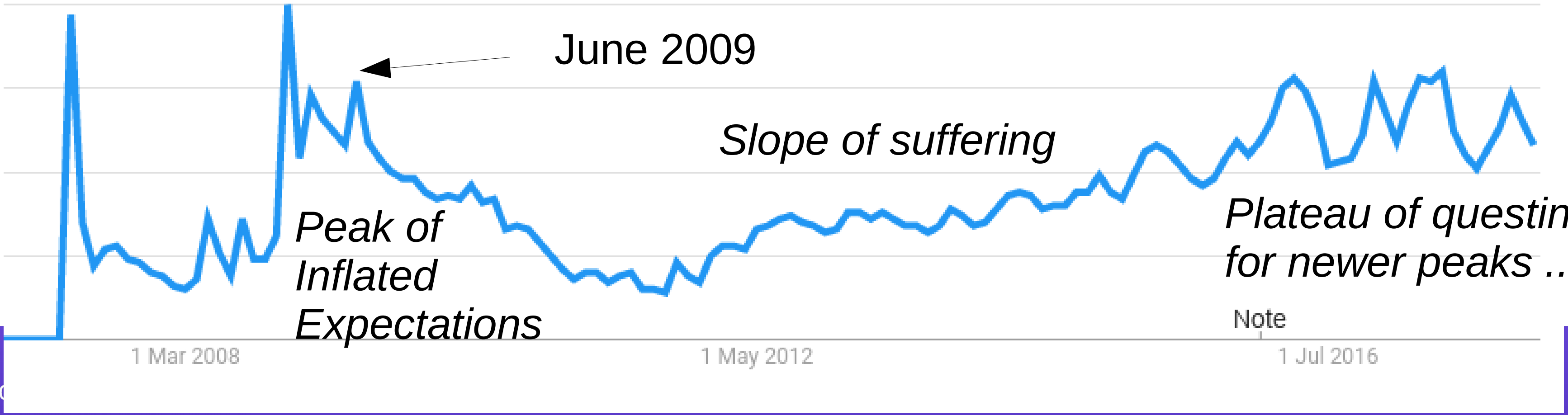
Sample 10x productivity wins

- Object Oriented programming !
- Java → develop 10x faster
  - Garbage collection ! It rocks.
- C# → develop 10x faster
  - Syntactic sugar! It rocks.
- "*As such, whilst Vala is a modern language with all of the features you would expect, it gains its power from …*"
  - Language augmenting pre-processors are great
- And on … and on.

A personal favourite

“We encourage the OpenOffice group to quickly build their version of a spread sheet or a word app using JavaFX," Ellison (according to theregister) - June 2009

- Fatal mistake: using a unique name
  - **Tip#2** – your vanity language / platform project should have a non-google-trend-able name eg. ‘ruby’, ‘rust’, ‘go’ - not eg. ‘zsquat’
- JavaFX’s Classic Hype-Cycle trajectory / google trend


June 2009

Peak of Inflated Expectations

Slope of suffering

Plateau of questing for newer peaks ...

Lucking out on language ?

"The point I'm trying to make is that the only reason why all of this was possible, where we build … [Microsoft Office] … for ***30 years****, not only ship those applications but evolve them - picked C/C++ - tremendously lucked out. All the other languages - two categories:*

1. language would have been ***re-invented*** *3 times and had to* ***re-write*** *the thing 3 times.*

2. authors of language & industry ***lost interest*** *and it would have* ***stagnated***"

– Igor Zaika (Microsoft, Distinguished Eng.)

CppCon 2014

Language bits.

- Language bigotry:
  - Almost totally pointless
  - **The** leading driver of pointless duplication
    - Over even license choice … ?
  - Lots of it out there.
- We’re still slowly writing out Java
  - can’t guerentee it is on the platform
    in the same way that binaries are.
  - The embedded Rhino JS-in-Java impl ? …
  - Default HSQLDB format → in databases …

Cross Platform-ness: Churn

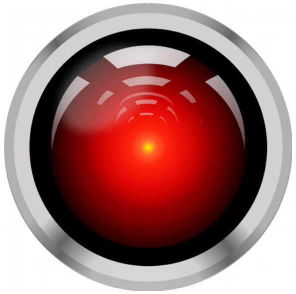
- Windows rendering APIs:
  - **GDI:** basic version with windows 1.0~1985
    - Made the 16bit → 64bit transition well.
- GDI+ in XP - 2001
- **DirectWrite** Win7/2007
  - finally a physical font API.
- Direct2D: 2012
- Easy to churn faster than big apps can handle

- Linux toolkits
  - gtk+1, Qt2 – 1999
  - **gtk+2**, Qt3 – 2000
  - **Qt4** 2005
  - **gtk+3** 2011
  - *Qt5* 2012
- Apple: churn to match
- Bold: in-use ...

Form factor changes

- Which changes to sit-out ?
- PC → mobile → netbook → tablet → web/browser → watch → VR → audio assistant: nice ...

  “I’m afraid I can’t save your document Dave”

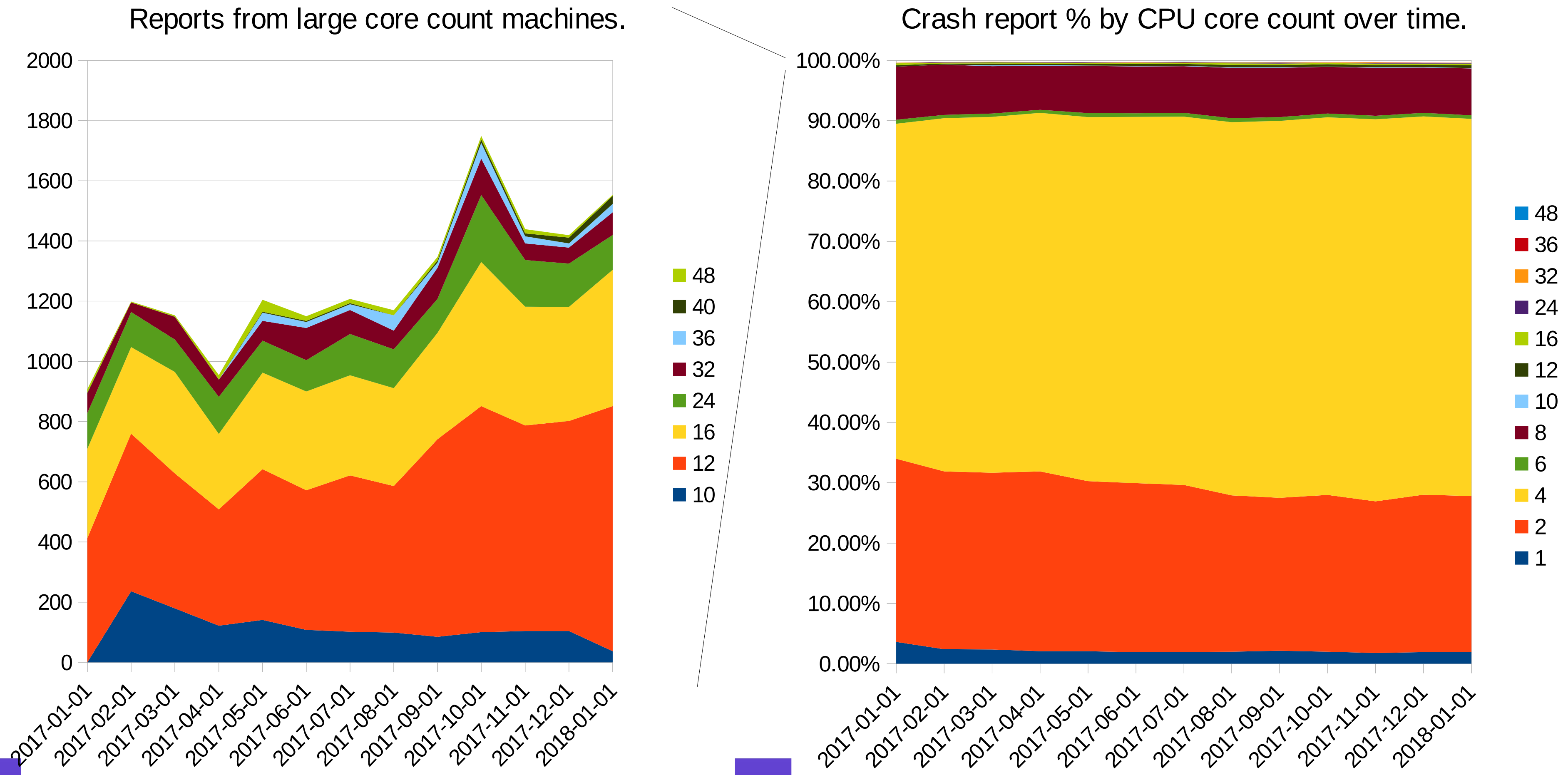


- More trivially:
  - Threads – the world is filling with them thanks to AMD:

RYZEN
THREADRIPPER

2017 Crash reporting stats

- Frustratingly ‘cores’ not threads.



Thanks to Markus Mohrhard

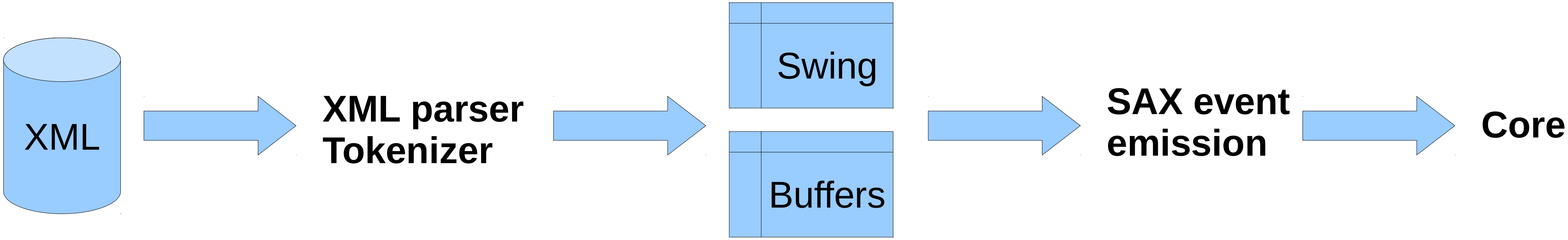
Threading

Newish Threading ...

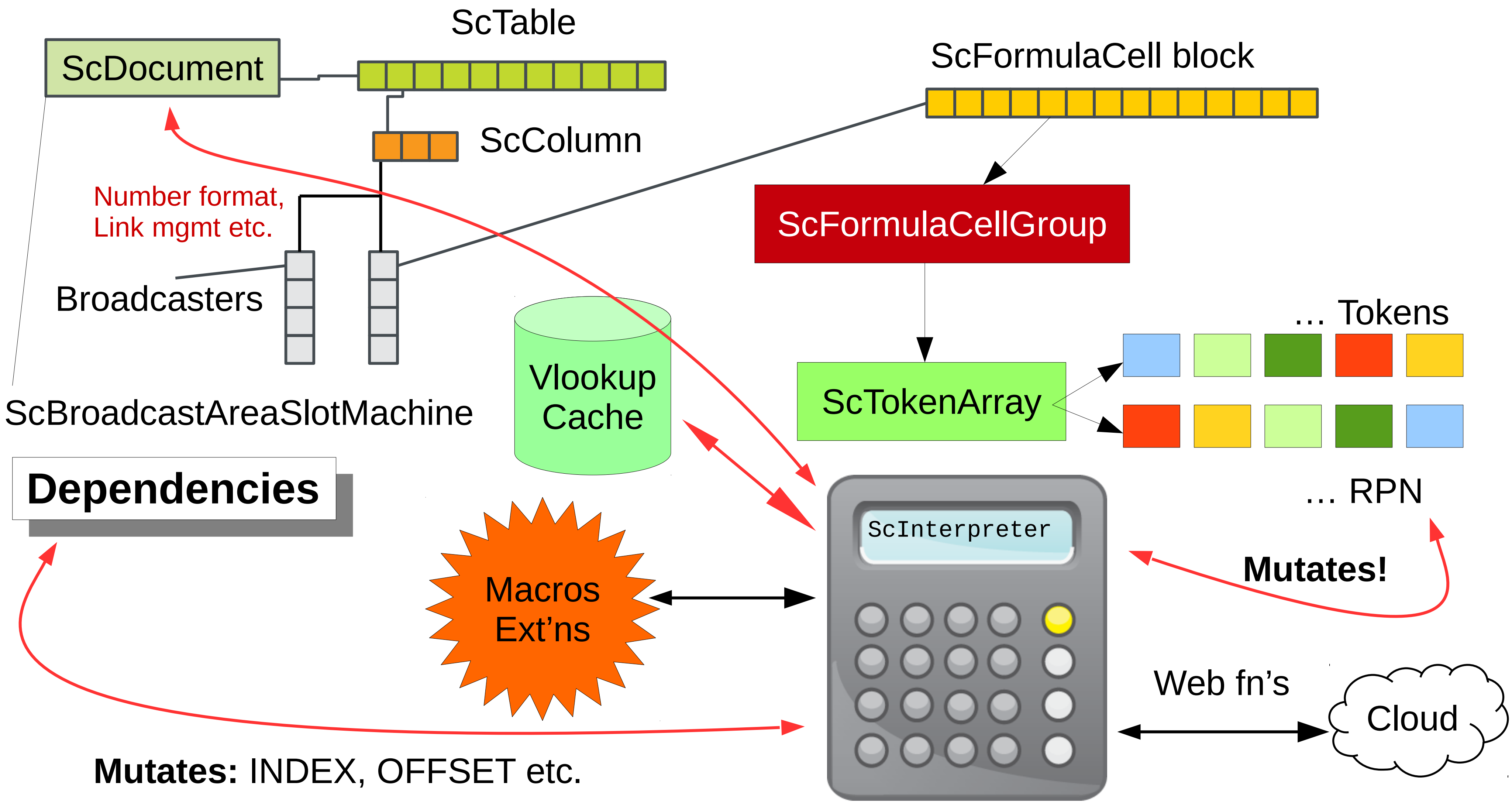
- CPU / Image scaling
- Rendering / Rasterizing primitives
  - *Thanks to Armin Le Grand (CIB)*
- Calc core
  - *Thanks to Tor Lillqvist & Dennis Francis (Collabora)*
- XML parsing: expansion to ODF filters
  - Constant time XML parsing (with free threads, and ~slower consumer)
  - *Thanks to Mohammed Abdul Azeem*

Threaded XML parsing ...

- Parsing XML is a significantly costly pain
  - Format is needlessly complex; cf. JSON.
  - Namespace handling & tokenization is a pain.
  - Double checking for duplicate attributes bad too ...
  - SAX API really poor for CPU cache usage:
    - a heavily fragmented workload.

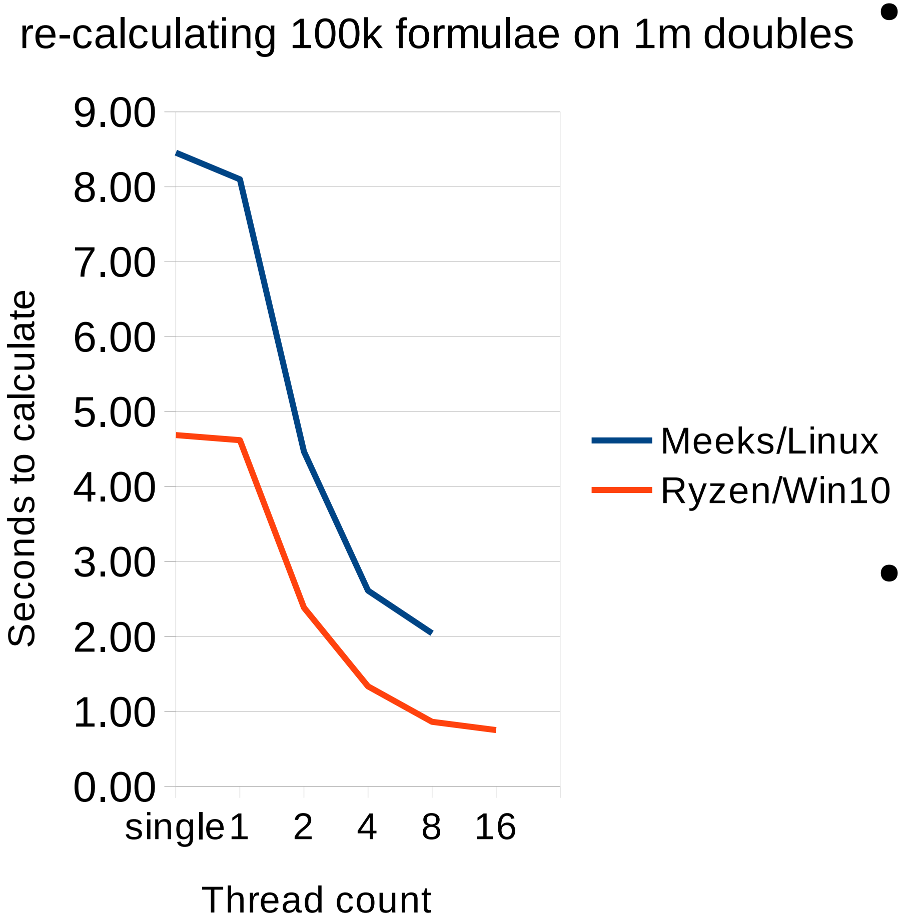


Calc: extreme coupling & threads.



Mutates: INDEX, OFFSET etc.

Calc parallelism results ...



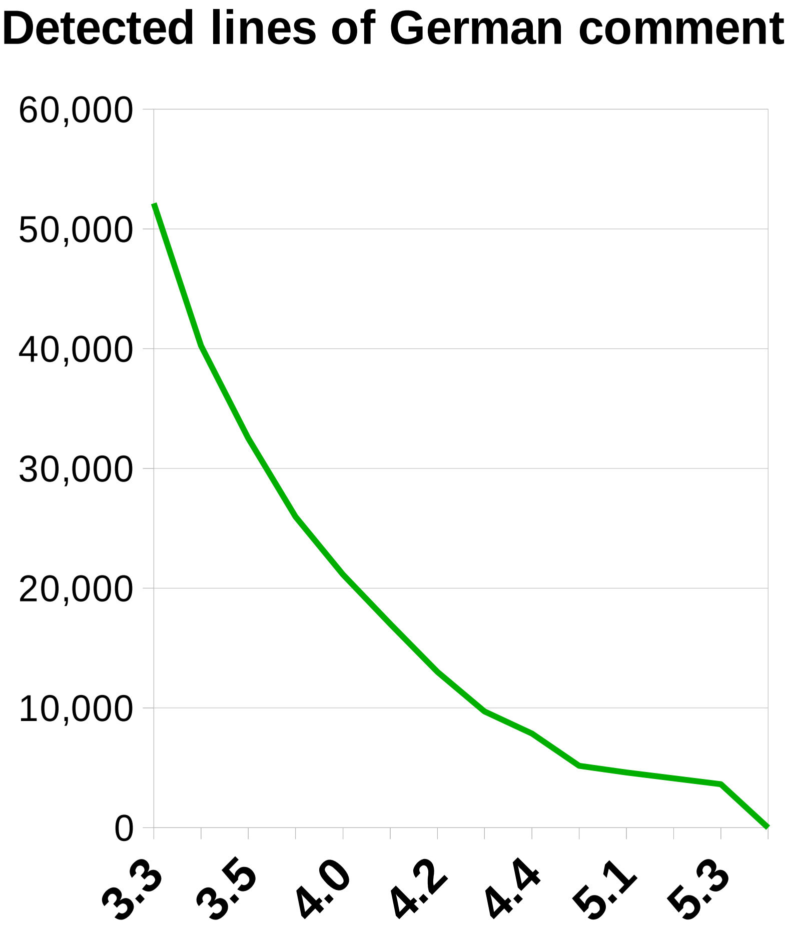
- Extreme care & choice of threading granularity
  - Mostly lock-less, but lots of assertions.
  - Would love more language support to help; C++ weak here.
- Some good improvements
  - Plenty more to do to expand the scope & reliability

Cleanups ...

Tastier code – easier to chew

- Finally finished quality translation in 5.4 of all German comments.
- Thanks to:

  Jens Carl, Johnny_M, Michael Stahl (RedHat), Katerina Behrens (CIB), Thomas Beck (CIB), Lukas Röllin, Gabor Kelemen, Stephan Bergmann (RedHat), dennisroczek, and many others.



Less eye-strain horror

- Ongoing enum cleanup & strings ...

```
-case BUTTON_ABORT: aText = rtl::OUString(
-      RTL_CONSTASCII_USTRINGPARAM( "Abort" ) );

+case StandardButtonType::Abort: aText = "Abort";
```

- Pretty iterators & auto

```
- for( std::unordered_map< Atom, Selection*
>::iterator it = This->m_aSelections.begin(); it !=
This->m_aSelections.end(); ++it )

+ for (auto const& selection : This→m_aSelections)
```

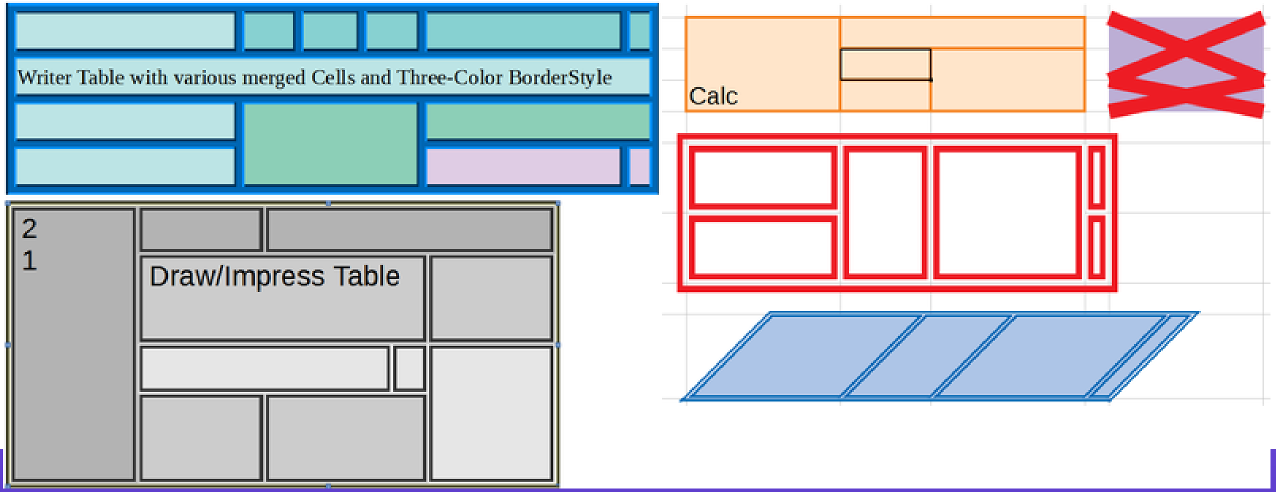
- And lots more.

Clang plugins ...

- Expanding C++ checking:
  - unexpected bool conversion hunting
  - flatten: look for large if statement at end of function
  - cast cleanups / re-writing
- Wider scope of understanding
  - Unused-fields checking
  - count all users of default params
  - un-necessary-virtual detection
  - Use-unique-ptr
- And tons more: thanks to
  - *Noel Grandin (Collabora) & Stephan Bergman (RedHat)*

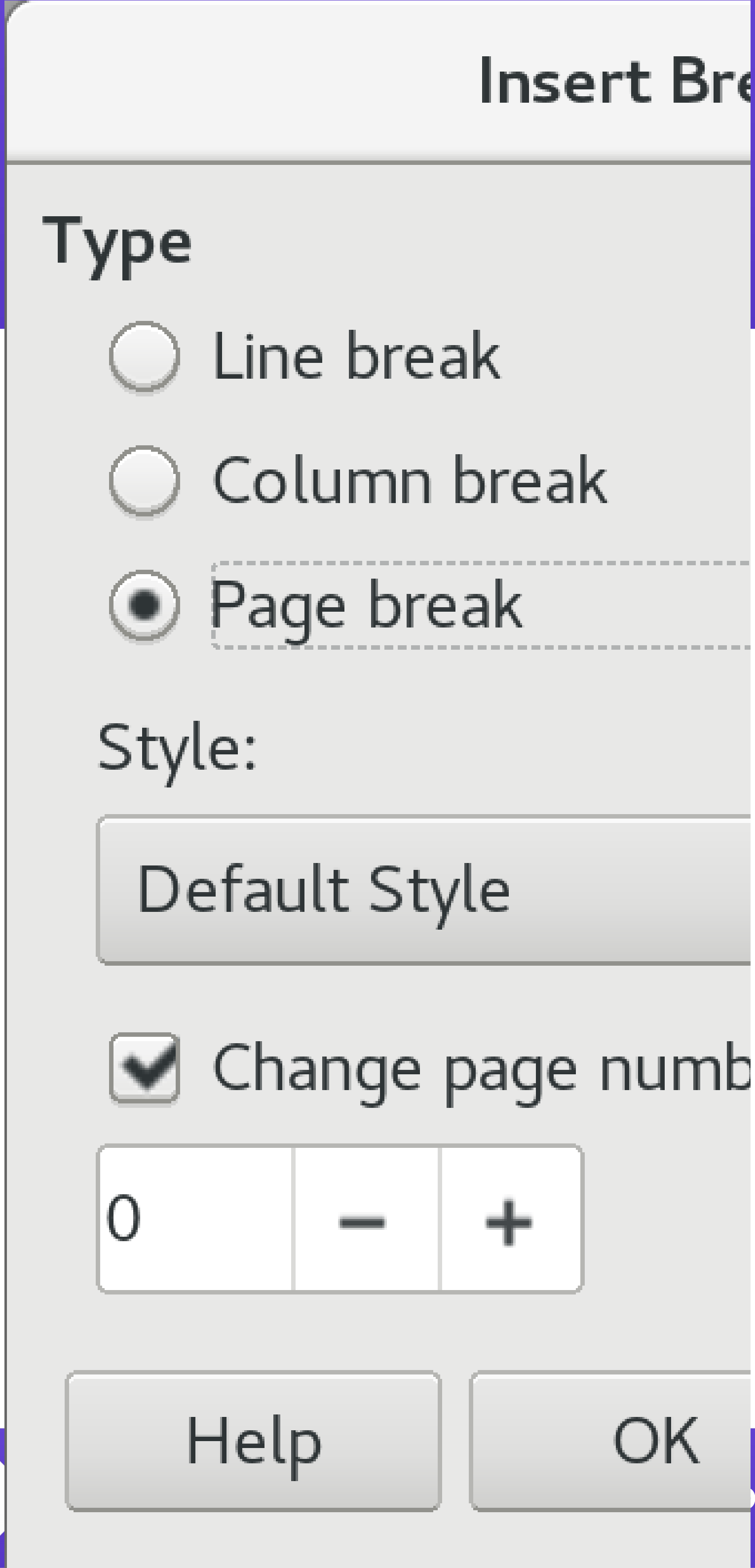
Drawing & Widgets ...

- Harfbuzz
  - Kills Uniscribe (Win), and CoreText (Apple)
  - Single, cross-platform, FLOSS shaping
    - Thanks to *Khaled Hosny & TDF donors*
  - Share glyph rendering too next ? Print...
- Drawinglayer Primitives for borders:
  - *Thanks to Armin Le Grand (CIB)*


Writer Table with various merged Cells and Three-Color BorderStyle
Calc
2
1
Draw/Impress Table

Gtk3: Native Widgets ...

- Native Widget theming
  - Used for years on Win/Mac/Linux
  - Torture widgets into rendering bits of themselves
  - Missing: theming details, animations
- Plus Native Menu-bar, File-selector, Tool-tips.
- Finished converting all dialogs to glade / XML years back: **977** .ui files ...
- Now starting to load native gtk+ dialogs, widgets & wrapping API – WxWidgets style
  - *Thanks to Caolan McNamara (RedHat)*



Insert Bre

Type

○ Line break

○ Column break

● Page break

Style:

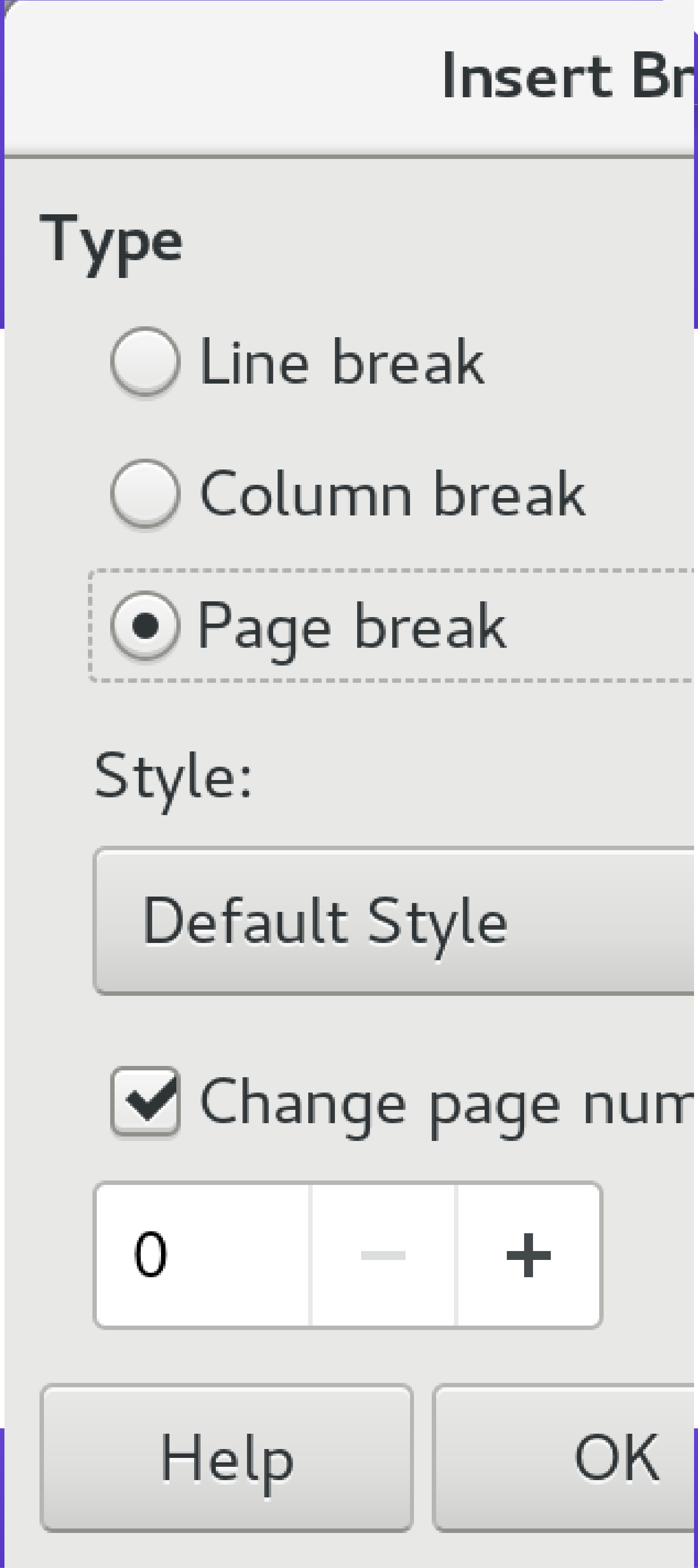
Default Style

☑ Change page numb

0 − +

Help OK

Video



Insert Br

Type

○ Line break

○ Column break

● Page break

Style:

Default Style

☑ Change page num

0 − +

Help OK

Core Infrastructure Initiative

THE LINUX FOUNDATION

- OSS-Fuzz
  - Huge Testing infrastructure provided by Google
    - Used for Chrome & many other OSS projects.
    - ~1000 core cluster to hugely accelerate testing.
  - Manual auditing – killed by AI ?
    - Finds newly introduced problems rapidly before they can escape (CVEs)
- Coverity – still a score of ~0.00
- *Thanks to Caolan McNamara (RedHat)*

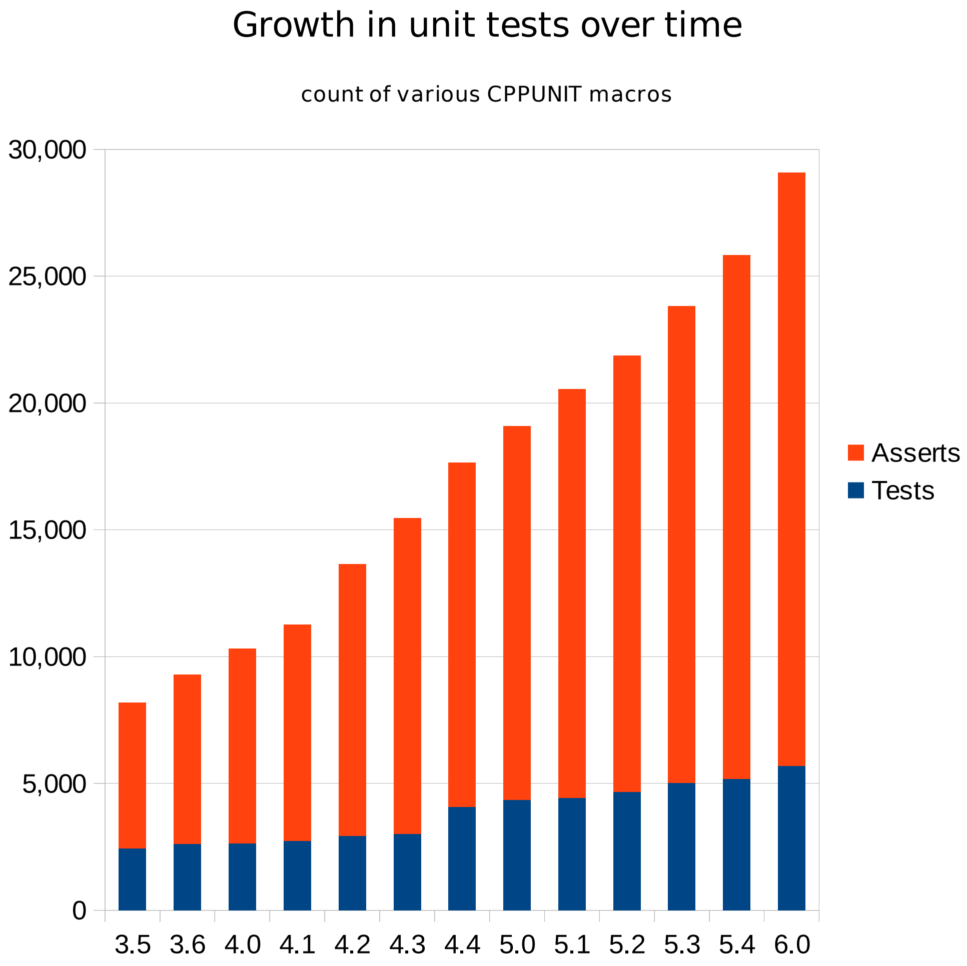
Quality Impact ?

Avoiding deja-vu in bug-fixing

- Fix each bug just once;

  Thanks to: Miklos Vajna (Collabora), Stephan Bergmann (RedHat), Noel Grandin (Collabora), Zdeněk Crhonek, Caolán McNamara (RedHat), Ashod Nakashian (Collabora), Justin Luth (SIL), Tamás Zolnai (Collabora), Andrea Gelmini, Jan Holesovsky (Collabora), Markus Mohrhard, Eike Rathke (RedHat), Mike Kaganski (Collabora), Jens Carl, Michael Stahl (RedHat), Szymon Kłos (Collabora), Tor Lillqvist (Collabora), Chris Sherlock, Pranav Kant (Collabora), Winfried Donkers, David Tardon (RedHat)

  with >20 unit test commits in 5.4/6.0



Re-factor to cause regressions ?

Commits:

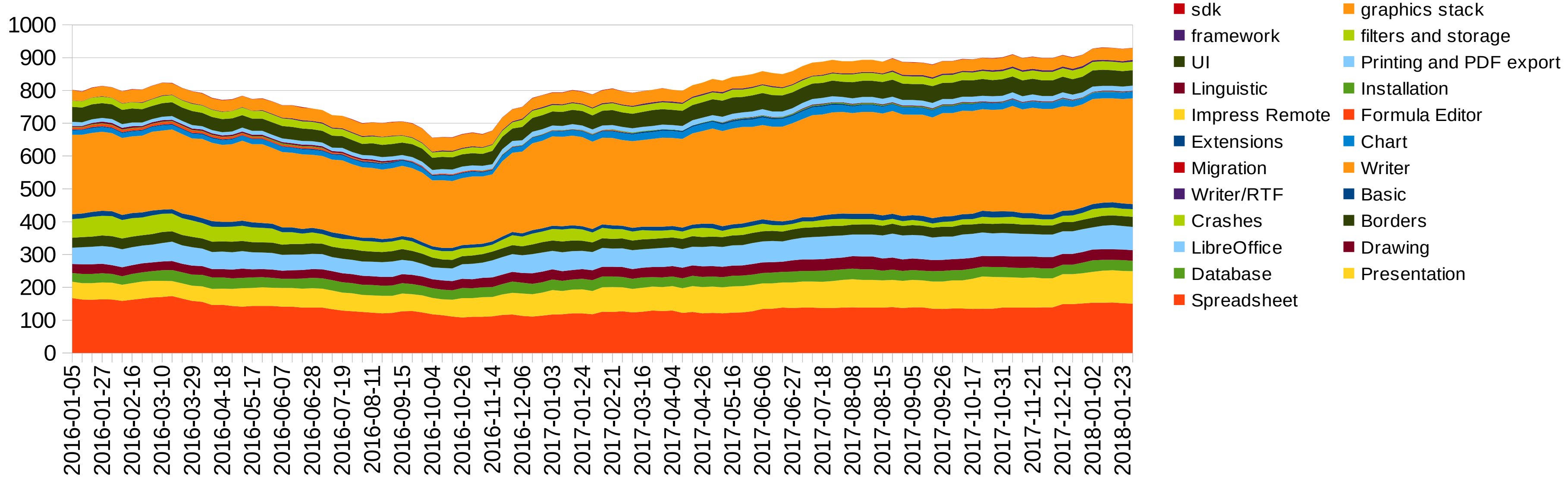
```
$ git log  --oneline –since ‘2016-01-01' | wc -l
33,040
```

Open regressions: +142 (0.4% of commits)

Not for free: lots of work **+2000** closed regressions in that time

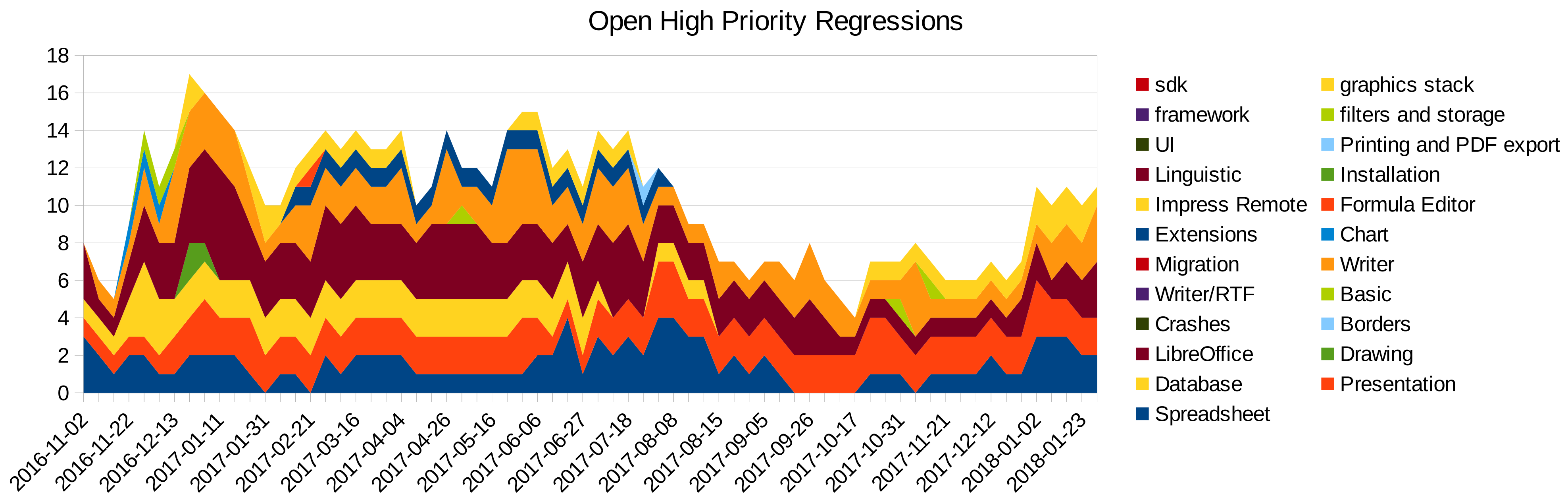
Many of which didn’t escape – *thanks to the QA team*

Open regression bugs 2016 - today ...

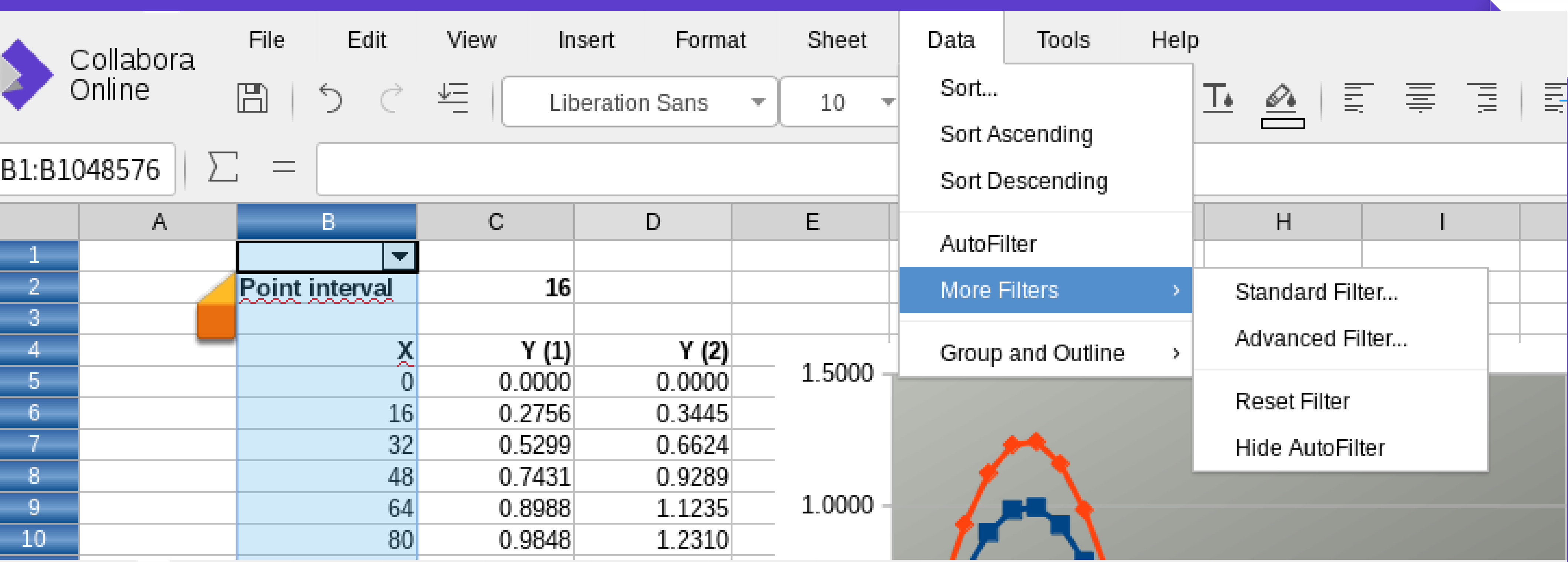


And how serious are they ?

- Under 10 high priority ones
  - Not escaping.
- Some bug fixes: rob Peter to pay Paul.
  - guarenteed: change a CPU/memory trade-off: bingo.

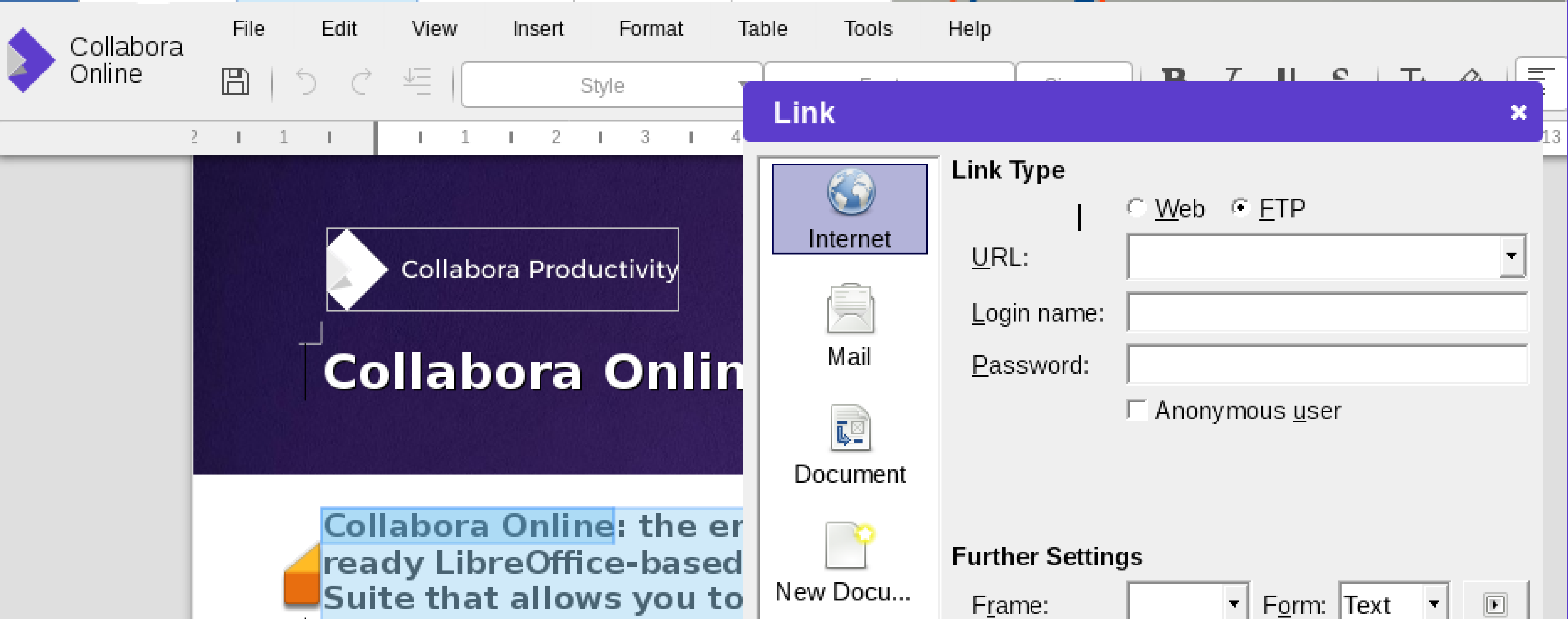

Open High Priority Regressions

Online – a rather different platform



B1:B1048576

	A	B	C	D
1				
2		Point interval	16	
3				
4		X	Y (1)	Y (2)
5		0	0.0000	0.0000
6		16	0.2756	0.3445
7		32	0.5299	0.6624
8		48	0.7431	0.9289
9		64	0.8988	1.1235
10		80	0.9848	1.2310



Collabora Productivity

Collabora Onlin

Collabora Online: the er
ready LibreOffice-based
Suite that allows you to

Link Type ○ Web ● FTP

URL:

Login name:

Password:

□ Anonymous user

Further Settings

Frame:

Form: Text

Different optimizations

- Online
  - dlopens ~world (RTLD_BIND_NOW)
  - Loads ~all dictionaries & hyphenation
  - then forks children
    - Copy-on-write-ness everywhere.
  - waste memory in pre-init to save in children
  - Waste CPU time on startup to save it later.
- Catching bad behaviour:
  - Touching / un-sharing pages:

Compare memory: loolmap

- Useful tool for eg. kdeinit, webservers?

```
heap page: 0x011e9000 (310/967) - touched: 17 - was shared - from [heap]
...
0x0130  03 00 00 00 0a 00 00 00  4d 00 61 00 74 00 68 00  | ........M.a.t.h.
-par't-  1                                                |
0x0140  4d 00 4c 00 20 00 32 00  2e 00 30 00 00 00 65 00  | M.L. .2...0...e.
0x0150  00 00 00 00 00 00 00 00  61 00 00 00 00 00 00 00  | ........a.......
```

Ref-count in a Unicode string incremented

- Also dumps all strings it can find with –strings
- Totals for heap

  -
    ```
    shared   23040 (92160kB)
    unshared   982 (3928kB)
    same but unshared       191 (764kB)
    dirtied bytes touched  30718 per page 31.28
    ```

Drove pre-init string staticization

Memory corner cases:

```
pixman/pixman-glyph.c (pixman_glyph_insert) - 40Mb

/* XXX: These numbers are arbitrary---we've never done any
measurements.

#define N_GLYPHS_HIGH_WATER  (16384)

#define N_GLYPHS_LOW_WATER   (8192)
```

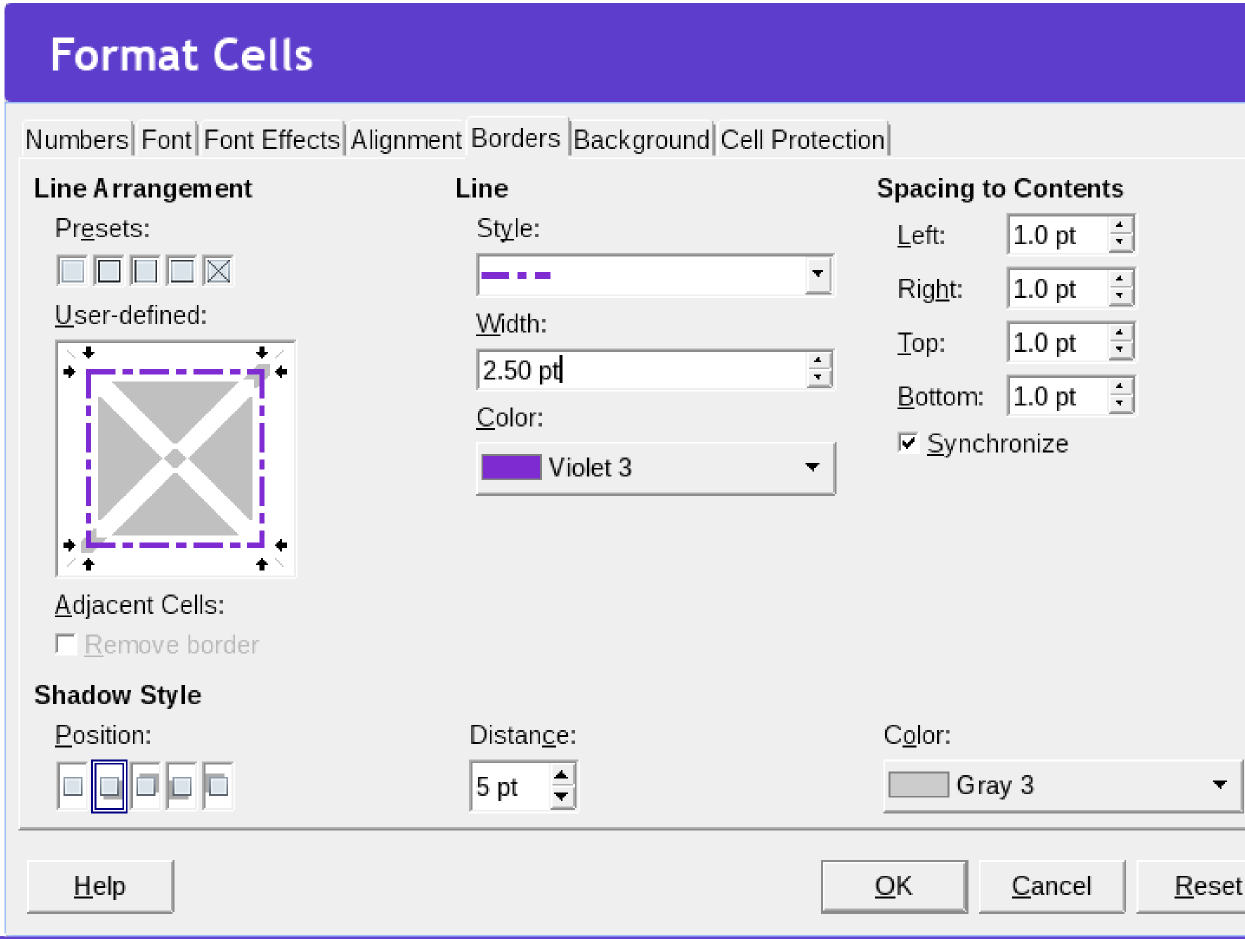
- How many glyph bitmaps to keep around.

```
cairo/src/cairo-scaled-font.c - likewise

/* XXX: This number is arbitrary---we've never done any
measurement of this. */

#define MAX_GLYPH_PAGES_CACHED 512
```

- Need control: ideally pre-render common glyphs patching pixman / cairo ? ...

Tunnelling dialogs ...

- Encouraging progress here for 6.1
- Rich dialogs
  - Deep function.
- Collaborative:
  - A. Edit border
  - B. Edit color
  - C. Edit size
- Applies right.
- Modal dialogs



Format Cells

Numbers | Font | Font Effects | Alignment | Borders | Background | Cell Protection

Line Arrangement

Presets:

User-defined:

Adjacent Cells:

□ Remove border

Line

Style:

Width: 2.50 pt

Color: Violet 3

Spacing to Contents

Left: 1.0 pt

Right: 1.0 pt

Top: 1.0 pt

Bottom: 1.0 pt

☑ Synchronize

Shadow Style

Position:

Distance: 5 pt

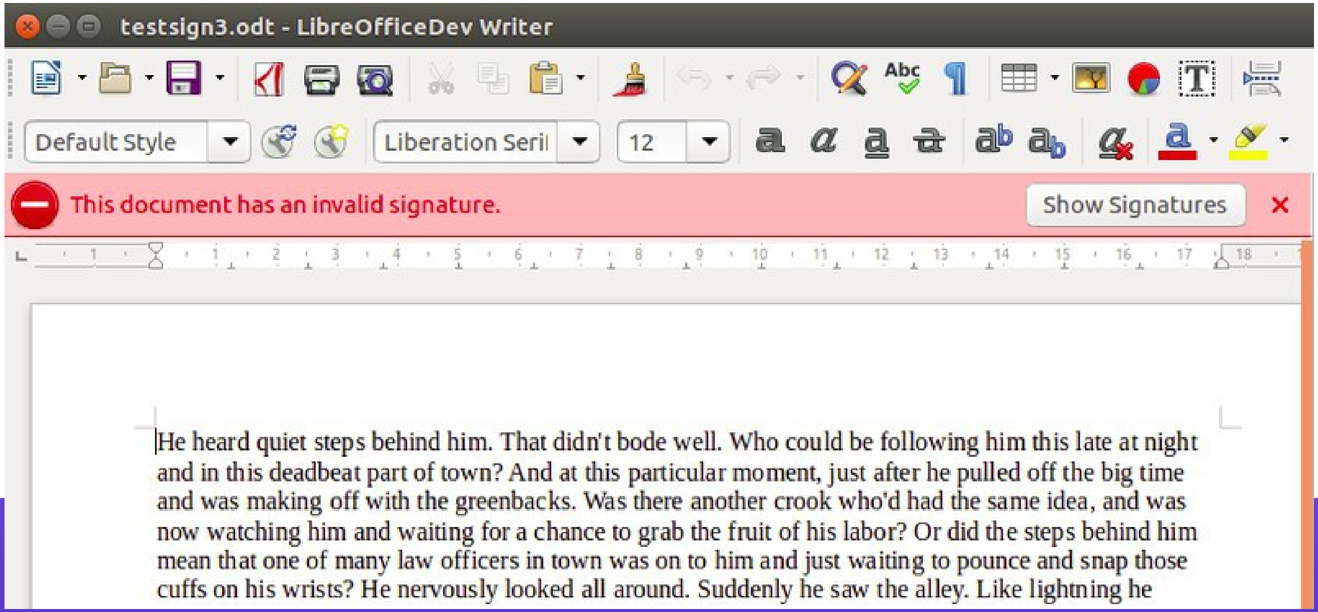
Color: Gray 3

Help | OK | Cancel | Reset

Small selection of 6.0 new features

Documents

- Mail merge from writer tables … *(Miklos Vajna – Collabora)*
- Image rotation *(Armin Le Grand – CIB)*
- EPUB3 export filter *(Miklos Vajna)*
- OpenPGP document sign & encrypt *(Katarína Behrens, Samuel Mehrbrodt, Thorsten Behrens - CIB)*

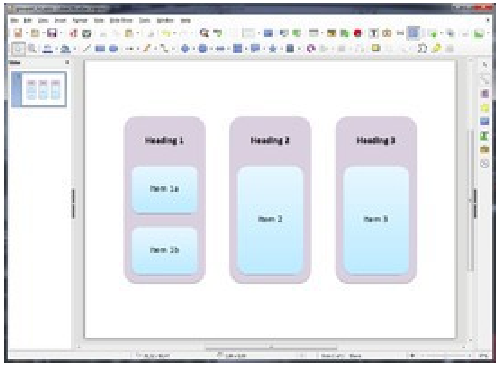


This document has an invalid signature.

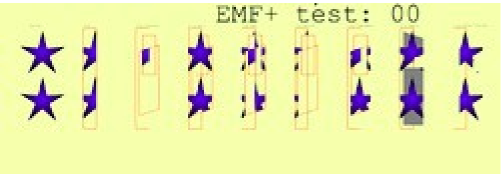
He heard quiet steps behind him. That didn't bode well. Who could be following him this late at night and in this deadbeat part of town? And at this particular moment, just after he pulled off the big time and was making off with the greenbacks. Was there another crook who'd had the same idea, and was now watching him and waiting for a chance to grab the fruit of his labor? Or did the steps behind him mean that one of many law officers in town was on to him and just waiting to pounce and snap those cuffs on his wrists? He nervously looked all around. Suddenly he saw the alley. Like lightning he

Filters

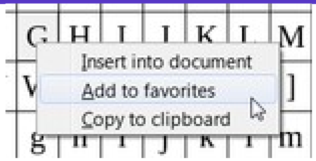
- QuarkXPress v3 & 4 import via libqxp from DLP *(Alex Pantechovskis)*
- Improved SmartArt import filter *(Grzegorz Araminowicz)*



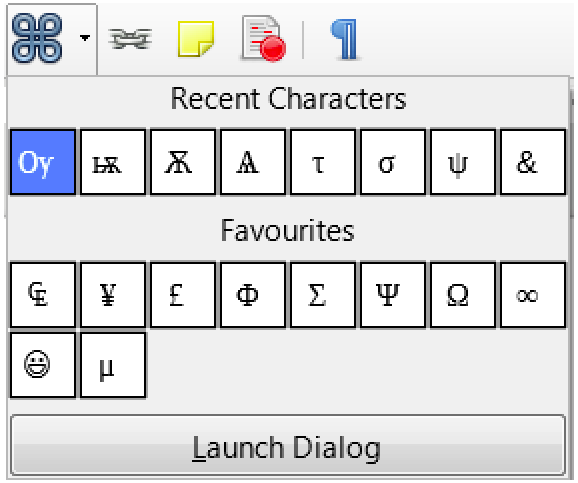
- ActiveX import/export to/from DOCX *(Tamas Zolnai – Collabora)*
- PPTX – export embedded videos, and round-trip macros *(Kendy – Collabora)*
- Many EMF+ filter fixes *(Bartosz Kosiorek¸Patrick Jaap)*



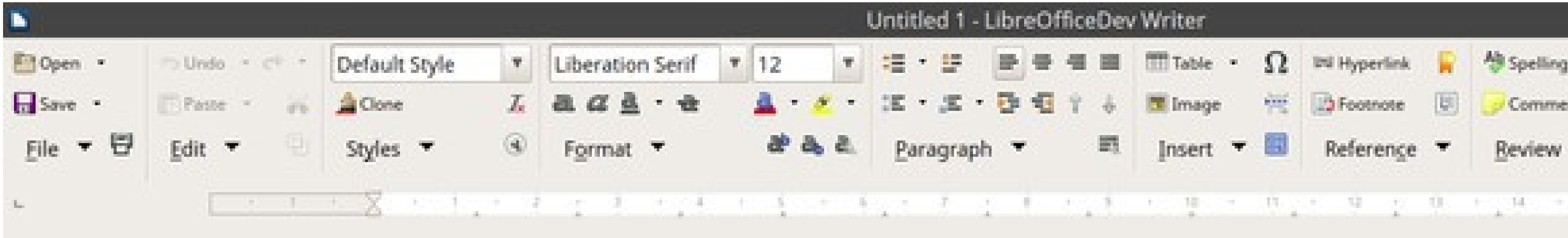
Quick GUI selection ...



- Faster Insert Special Characters *(Akshay Deep)*

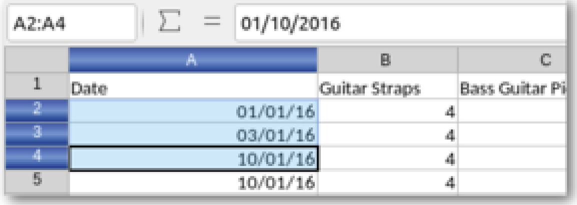


- Improved Notebookbar *(Andreas Kainz)*



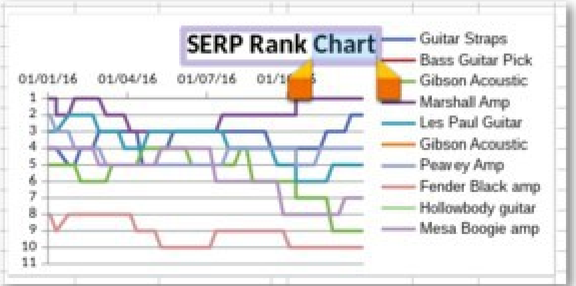
Other bits

- We dropped support for Windows XP
  - And Windows Vista … at last.
- Online *(from Collaborans)*
  - Seccomp-bpf, Avatars, Spell-checking
  - Interactive horizontal ruler
  - Calc rows → 500k.
  - Calc: active cell reference
  - Basic chart editing …



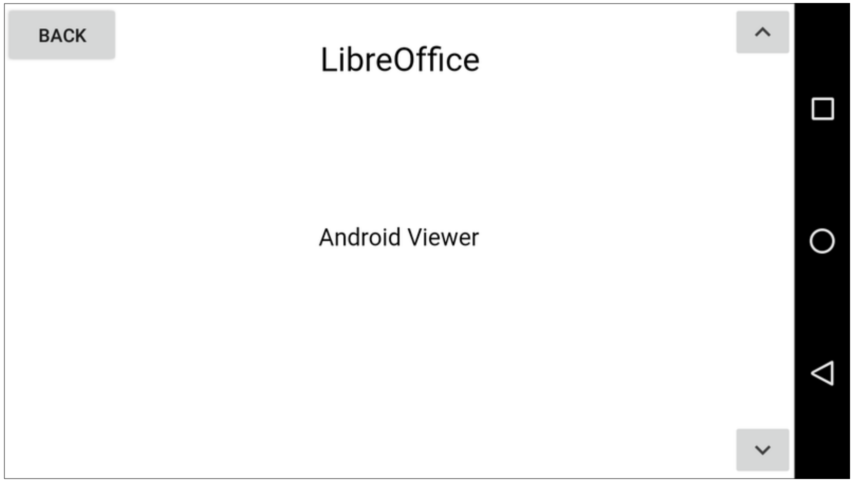
A2:A4 ∑ = 01/10/2016

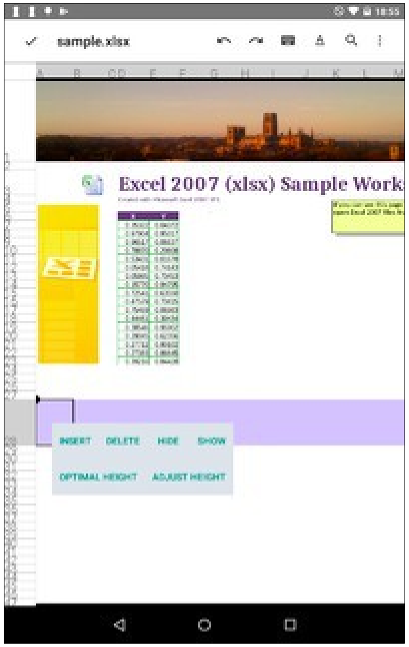
	A	B	C
1	Date	Guitar Straps	Bass Guitar Pi
2	01/01/16	4	
3	03/01/16	4	
4	10/01/16	4	
5	10/01/16	4	

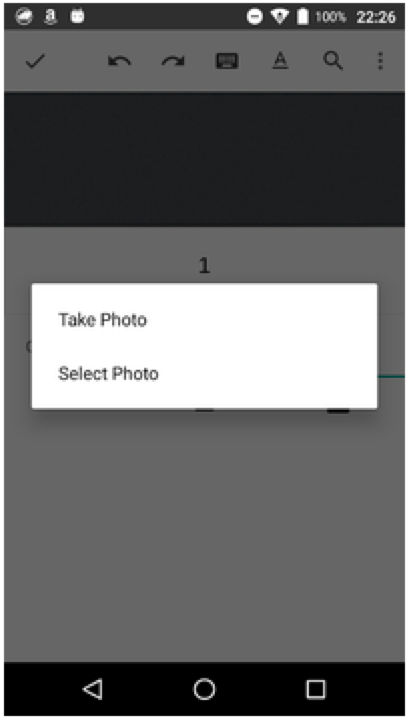


Android version *(Ximeng Zu)*

- Insert pictures from camera:
- Configure rows:
  - Insert, delete, hide
- Presentation mode







IOS progress: *thanks to Jan Iversen & Jon Nermut*

Getting involved.

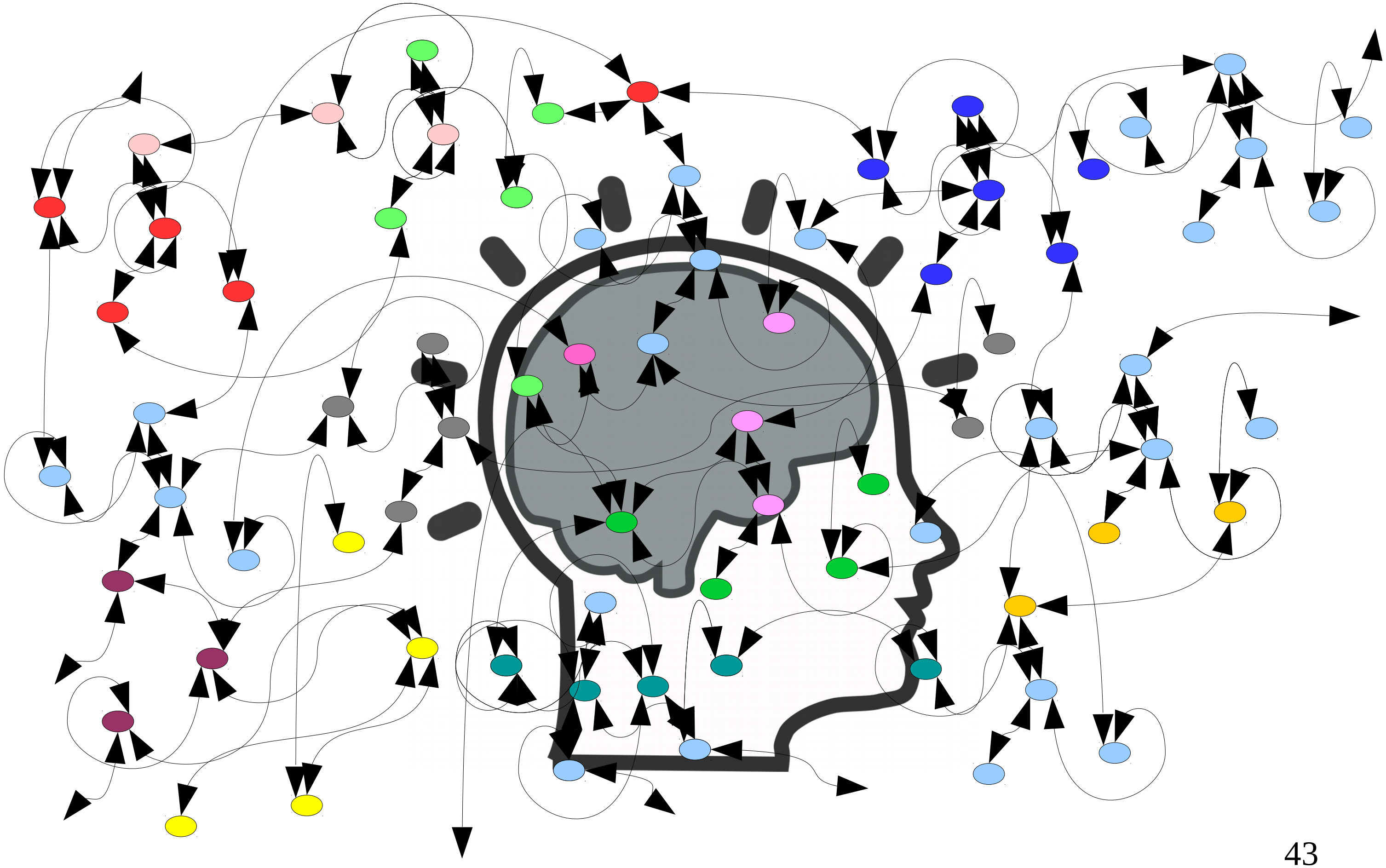
Getting involved

- Something for everyone:
  - Testing – just run the latest builds / triage
  - Documentation – help mend the manuals
  - Ask – see if you can answer a question.
  - Developing – poke at an easy hack
    - Python, Javascript, C++ tasks ...
    - From the very simplest to something ‘fun’
  - Design / Art – join a team call

https://www.libreoffice.org/community/get-involved/

Nurture scaleable development

- Try working a problem at doesn’t fit your head
- These are the ‘real’ engineering problems once you can solve them you’re a star.
- Checkout my paper on the topic.



The future ...

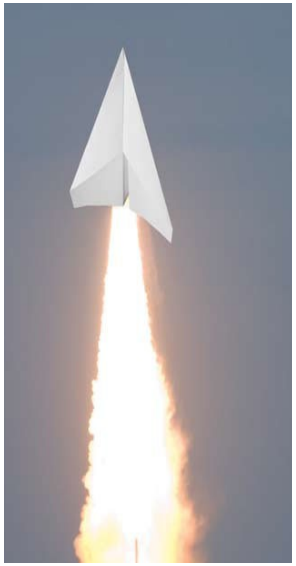
- We need smart people to help us
- Create the future:
  - An Artificial Intelligence

    of stunning subtlety

    of conversational brilliance
  - That can understand what you’re doing
  - And can suggest cool new things to do.
  - Trained on Documents (TM)

Conclusions

- Forced platform change is tough on big code-bases
- Re-structuring & re-factoring
  - fun – and risky: today seems hard to justify.
  - stopping re-factoring is also risky.
- LibreOffice is surviving the volume of change & improving
- Online – is great, free your data.
- New contributors are always welcome : get stuck in
- Thanks for supporting LibreOffice !



Oh, that my words were recorded, that they were written on a scroll, that they were inscribed with an iron tool on lead, or engraved in rock for ever! I know that my Redeemer lives, and that in the end he will stand upon the earth. And though this body has been destroyed yet in my flesh I will see God, I myself will see him, with my own eyes - I and not another. How my heart yearns within me. - Job 19: 23-27