

Developments at NISZ

Tips & tricks behind the scenes

László Németh (nemeth@numbertext.org)

~170 bug fixes

- From July 2018 to September 2019
 - ~100 bug fixes by me (OOXML, track changes)
 - ~50 bug fixes by Balázs Varga (charts)
 - ~20 bug fixes by Tünde Tóth, Gábor Kelemen, Ádám Kovács, József Szakács, Szabolcs Tóth, Bálint Sass and László Zoltán Kovács
- Recent team: 4 members + 3 newcomers + a student trainee

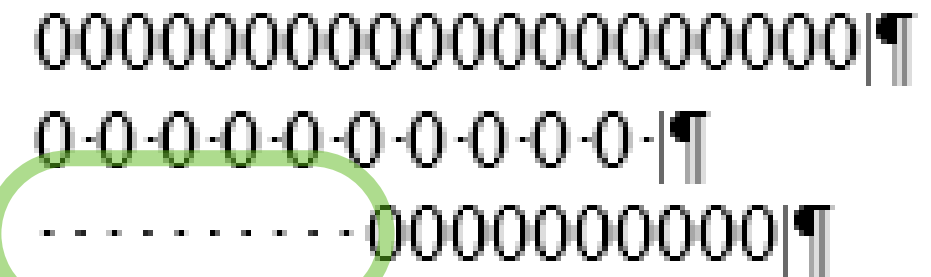
How to develop LibreOffice?

- Understand the problem
- Choose the right tool
- Read the code actively
- Fix the bug or find a workaround

Understand the problem


- In RTF documents, “size of spaces is different in Word and Writer”
- UI TESTING: only in space sequences!
- +1/6 em

Word



A hex dump showing the RTF space sequence for Microsoft Word. The first line contains 16 zeros followed by a paragraph mark (0x0D, 0x0A). The second line contains 10 pairs of '0' and '0' followed by a paragraph mark. The third line contains 10 dashes followed by 8 zeros and a paragraph mark. A green oval highlights the sequence of dashes and zeros in the third line.

Writer



A hex dump showing the RTF space sequence for LibreOffice Writer. The first line contains a paragraph mark (0x0D, 0x0A) followed by 16 zeros. The second line contains 10 pairs of '0' and '0' followed by a paragraph mark. The third line contains 10 dashes followed by 8 zeros and a paragraph mark. A red oval highlights the sequence of dashes and zeros in the third line.

Choose the right tool

- LO/MSO/Google Docs etc.:
 - UI (editing, settings, options)
 - Import tests
 - Export tests
- XML (OOXML or ODF) analysis
 - Create test pairs with a single difference
 - Compare them by odiff or Open XML SDK

http://numbertext.org/odiff

- odiff file1.docx file2.docx
- odiff file1.odt file2.odt
- odiff file1.odt # show XML content

```
<w:r>  
  <w:rPr/>  
  <w:t>Lorem ipsum dolor sit amet, consectetur adipiscing elit. elit.</w:t>
```

```
</w:r>  
<w:r>  
  <w:rPr/>  
  <w:commentReference w:id="0"/>
```

```
</w:r>  
<w:r>  
  <w:rPr/>  
  <w:t xml:space="preserve"> Vestibulum consequat mi quis pretium semper. Proin  
luctus orci ac neque venenatis, quis commodo dolor posuere. Curabitur dignissim sapien  
quis cursus egestas. Donec blandit auctor arcu, nec pellentesque eros molestie eget.  
In consectetur aliquam hendrerit. Sed cursus mauris vitae ligula pellentesque, non pel  
lentesque urna aliquet. Fusce placerat mauris enim, nec rutrum purus semper vel. Praes
```

Read the code *actively*

- git grep commentReference
 sw/source/filter/ww8/docxattributeoutput.cxx:
 m_pSerializer->singleElementNS(XML_w,
 XML_commentReference, FSNS(XML_w, XML_id
), OString::number(nId));
- SAL_DEBUG("Comment ID: " << nId);
- soffice --convert-to docx comment.odt
 debug:16080:16080: Comment ID: 0

Fix the bug

- Code reuse
 - git grep, git log
- Keep it stupid simple (KISS)
- Break the problem to a few parts
 - https://en.wikipedia.org/wiki/Pareto_principle#In_computing
- Create unit tests (copy and modify)

Unit tests (sw/qa, sc/qa...)

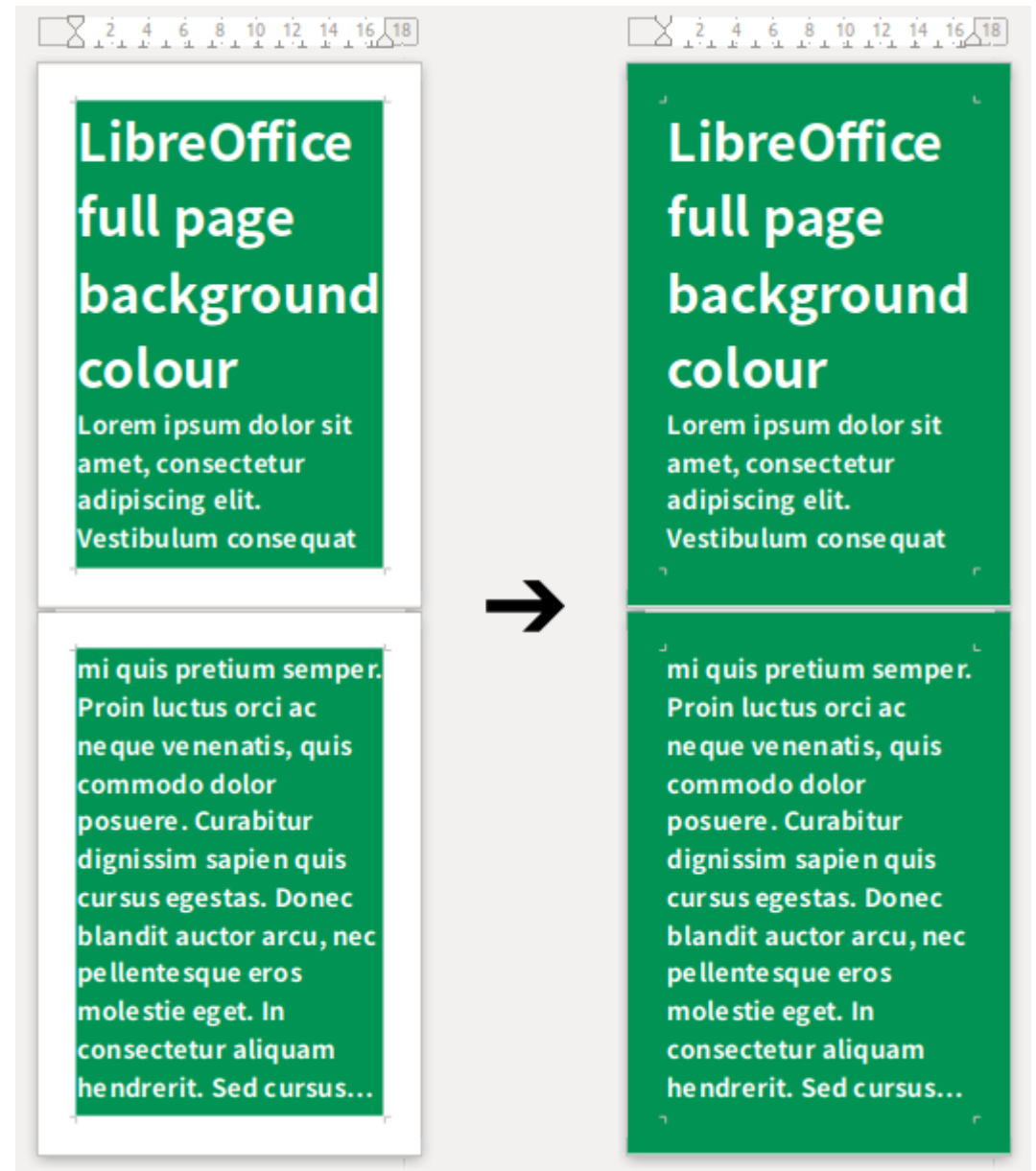
- XPath (document format)
- Inner API
- UNO API
- GDIMetaFile (page layout as an image)
- UI (editing, ie. uiwriter)
- UI dialog windows/Python

Speed up development

- Growth mindset
- Continuous mentoring
- Use gerrit and (libreoffice-dev or company level) developer channels

Example

- Colored pages in Writer



Solution

```
diff --git a/sw/source/core/layout/paintfrm.cxx b/sw/source/core/layout/paintfrm.cxx
index 1adf941..e920046 100644
--- a/sw/source/core/layout/paintfrm.cxx
+++ b/sw/source/core/layout/paintfrm.cxx
@@ -6179,7 +6179,15 @@ void SwFrame::PaintSwFrameBackground( const SwRect &rRect, const SwPageFrame *pP
    }
    else
    {
-        ::lcl_CalcBorderRect( aRect, this, rAttrs, false, gProp);
+        if ( bPageFrame )
+        {
+            aRect = getFrameArea();
+        }
+        else
+        {
+            ::lcl_CalcBorderRect( aRect, this, rAttrs, false, gProp);
+        }

        if ( (IsTextFrame() || IsTabFrame()) && GetPrev() )
        {
            if ( GetPrev()->GetAttrSet()->GetBackground() == GetAttrSet()->GetBackground() &&
```

Example

**Copy filtered
Calc data into
Writer tables**

	A	B
1	name	sex
2	Aaliyah	female
5	Abby	female
9	Abigail	female
14	Ada	female
15	Adalee	female
16	Adaline	female



Before 6.2

	Aaliyah	female
	Abby	female
	Abigail	female
	Ada	female

6.2

Aaliyah	female
Aarav	male
Aaron	male
Abby	female
Abdiel	male
Abdullah	male
Abel	male
Abigail	female
Abraham	male
Abram	male
Ace	male
Achilles	male
Ada	female

6.3

Aaliyah	female
Abby	female
Abigail	female
Ada	female

Solution

```
@@ -1187,6 +1192,26 @@ bool SwTransferable::Paste(SwWrtShell& rSh, TransferableDataHelper& rData, RndSt
    }
}

+ // tdf#37223 insert OLE table in text tables as a native text table
+ // (first as an RTF nested table, and cut and paste that to get a
+ // native table insertion, removing also the temporary nested table)
+ // TODO set a working view lock to avoid of showing the temporary nested table for a moment
+ if (bInsertOleTableInTable && EXCHG_OUT_ACTION_INSERT_STRING == nAction)
+ {
+     bool bPasted = SwTransferable::PasteData( rData, rSh, nAction, nActionFlags, nFormat,
+                                               nDestination, false, false, nullptr, 0, false, nAnchorType );
+     if (bPasted && rSh.DoesUndo())
+     {
+         SfxDispatcher* pDispatch = rSh.GetView().GetViewFrame()->GetDispatcher();
+         pDispatch->Execute(FN_PREV_TABLE, SfxCallMode::SYNCHRON);
+         pDispatch->Execute(FN_TABLE_SELECT_ALL, SfxCallMode::SYNCHRON);
+         pDispatch->Execute(SID_COPY, SfxCallMode::SYNCHRON);
+         pDispatch->Execute(SID_UNDO, SfxCallMode::SYNCHRON);
+         pDispatch->Execute(SID_PASTE, SfxCallMode::SYNCHRON);
+     }
+     return bPasted;
+ }
+ }
```

Example

~~Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.~~

Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin pharetra nonurn

Aenean nec lora

Suspendisse du
sem venenatis e

Lorem ipsum d
posuere, magna
quis urna.

~~Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.~~

Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin pharetra nonummy pede. Mauris et orci.

Aenean nec lorem. In porttitor. Donec laoreet nonummy augue.

Suspendisse dui purus, scelerisque at, vulputate vitae, pretium mattis, nunc. Mauris eget neque at sem venenatis eleifend. Ut nonummy.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.



Solution

```
diff --git a/writerfilter/source/dmapper/DomainMapper_Impl.cxx
index 8a9c038..321866d 100644
--- a/writerfilter/source/dmapper/DomainMapper_Impl.cxx
+++ b/writerfilter/source/dmapper/DomainMapper_Impl.cxx
@@ -1860,9 +1860,11 @@ void DomainMapper_Impl::CreateRedline(un
    case XML_mod:
        sType = getPropertyName( PROP_FORMAT );
        break;
+   case XML_moveTo:
+   case XML_ins:
        sType = getPropertyName( PROP_INSERT );
        break;
+   case XML_moveFrom:
+   case XML_del:
        sType = getPropertyName( PROP_DELETE );
        break;
```


Example

Paragraph formatting changes ¶

Left-aligned (default) → now right-aligned (direct formatting) ¶

Centered (direct formatting) → now right-aligned (direct formatting) ¶

Centered (direct formatting) → now Left-aligned (default) ¶

Left-aligned with a modified region → now right-aligned with **modified regions** (direct formatting) ¶

Paragraph numbering changes ¶

Numbered → now not-numbered ¶

1. → **Not-numbered** → numbered ¶

Paragraph style changes ¶

Default text → now Heading 2 ¶

Heading 3 → now Heading 2 ¶

Paragraph formatting changes ¶

Left-aligned (default) → now right-aligned (direct formatting) ¶

Centered (direct formatting) → now right-aligned (direct formatting) ¶

Centered (direct formatting) → now Left-aligned (default) ¶

Left-aligned with a modified region → now right-aligned with modified regions (direct formatting) ¶

Paragraph numbering changes (without style changes) ¶

1. → **Numbered** → now not-numbered ¶

Not-numbered → numbered ¶

Paragraph style changes ¶

Default text → now Heading 2 ¶

Heading 3 → now Heading 2 ¶

12 hours	tdf#124491 DOCX: import tracked changes of empty paragraphs	László Németh	6	-5/+35
7 days	tdf#105485 DOCX: import deleted comments as deleted	László Németh	3	-1/+10
9 days	tdf#123703 strip six-em-space (U+2006) at line break	László Németh	4	-7/+9
9 days	tdf#123703 RTF import: fix length of space character sequence	László Németh	4	-1/+68
10 days	tdf#127101 Change tracking: fix format reject at paragraph join	László Németh	4	-6/+33
10 days	tdf#127101 Change tracking: reject format at paragraph join	László Németh	2	-17/+128
11 days	tdf#127097 Change tracking: change style of whole deletion at paragraph join	László Németh	2	-7/+46
2019-08-02	tdf#123702 RTF/DOCX default 1" left/right page margins	László Németh	2	-7/+7
2019-08-02	tdf#126243 DOCX change tracking: pStyle needs simplified style id	László Németh	4	-21/+28
2019-08-01	tdf#118699 change tracking: fix accept before empty line	László Németh	2	-53/+57
2019-08-01	tdf#118699 change tracking: don't number empty lines	László Németh	1	-1/+1
2019-08-01	tdf#125990 change tracking: remove text join workaround	László Németh	5	-92/+26
2019-07-23	tdf#126516 DOCX change tracking: fix paragraph style rejection	László Németh	4	-13/+44
2019-07-19	tdf#125546 DOCX import: fix overgrowth of change tracking entries	László Németh	5	-14/+39
2019-07-18	tdf#125894: DOCX: import tracked changes in frames	László Németh	5	-3/+59
2019-07-18	Revert "tdf#123627 DOCX import: fix relative hyperlinks to documents"	László Németh	3	-7/+6
2019-07-12	tdf#126245 DOCX export: fix losing of tracked numbering changes	László Németh	6	-1/+22
2019-07-12	tdf#126245 DOCX: reject tracked paragraph numbering changes	László Németh	3	-4/+43
2019-07-10	tdf#121663 DOCX import: fix missing space in line numbering	László Németh	4	-2/+16
2019-07-09	tdf#126243 DOCX: export/reject tracked paragraph style change	László Németh	7	-28/+59
2019-07-06	tdf#120338 DOCX: reject other paragraph formatting changes	László Németh	1	-6/+8
2019-07-05	tdf#120338 DOCX: add rejection of paragraph formatting changes	László Németh	3	-0/+60
2019-07-03	tdf#126206 DOCX: add rejection of character formatting changes	László Németh	4	-8/+48
2019-06-24	tdf#119571 fix style & numbering at tracked deletion	László Németh	2	-53/+131
2019-06-24	tdf#54819 keep style & numbering at tracked deletion	László Németh	4	-1/+129
2019-06-17	tdf#125916 DOCX import: fix outline numbering at tracked deletion	László Németh	3	-0/+19
2019-06-14	tdf#125881 DOCX import: handle list level after tracked deletion	László Németh	3	-4/+23
2019-06-13	tdf#89991 DOCX: import Show changes from older formats	László Németh	1	-2/+10
2019-06-12	tdf#89991 DOCX: import/export Show changes mode	László Németh	8	-5/+48
2019-06-07	sanitize LibreLogo calls	László Németh	1	-1/+50
2019-06-06	tdf#115212 DOCX export: fix ToC with tracked deletion	László Németh	3	-0/+9
2019-06-04	tdf#70234 DOCX export: fix only expanded fields	László Németh	1	-2/+2
2019-06-04	tdf#70234 DOCX: export tracked deletion of fields	László Németh	3	-6/+32
2019-06-04	tdf#70234 DOCX: import tracked deletion of fields	László Németh	3	-0/+8
2019-06-03	tdf#120412 char formatting UI: clean-up DFLT_ESC_AUTO	László Németh	4	-9/+21
2019-06-03	tdf#112567 XLSX export: correct built-in names	László Németh	3	-1/+55
2019-05-29	tdf#120412 character formatting UI: allow >100% raising	László Németh	1	-0/+3
2019-05-28	tdf#92335 DOCX: fix multiplying of "ListLabel" styles	László Németh	3	-0/+14
2019-05-28	DOCX import: clean-up FinalizeImport() of change tracking	László Németh	4	-4/+24
2019-05-27	tdf#115557 DOCX import: don't multiply charts in footnotes	László Németh	3	-9/+23
2019-05-24	tdf#125319 DOCX track changes: don't change numbering	László Németh	1	-3/+10
2019-05-22	tdf#121784 DOCX import: fix change tracking of footnotes	László Németh	3	-1/+21
2019-05-21	tdf#125310 change tracking: don't copy page break	László Németh	3	-1/+58
2019-05-20	tdf#125310 change tracking: don't copy page break	László Németh	3	-1/+64
2019-05-15	tdf#118699 DOCX import: don't add numbering	László Németh	4	-0/+56
2019-05-11	tdf#125187 DOCX track changes: fix w:del within w:ins	László Németh	2	-5/+5
2019-05-08	tdf#116501 fix freezing at embedded text tables	László Németh	6	-2/+44
2019-05-02	tdf#112567 XLSX export: fix broken built-in names	László Németh	3	-1/+37
2019-04-29	tdf#123460 DOCX track changes: moveFrom completely	László Németh	5	-8/+47
2019-04-26	tdf#67207 DOCX mail merge: fix export/import of database fields	László Németh	8	-2/+120

3 days	tdf#103988 OOXML export: fix bubble chart MSO interoperability	Balazs Varga	2	-3/+7
3 days	tdf#114657 OOXML chart export: fix broken chart with NaN X value	Balazs Varga	3	-18/+22
3 days	tdf#124083 OOXML import: fix broken charts with missing value X	Balazs Varga	3	-2/+27
11 days	tdf#127030 PPTX export: keep fixed size of legacy shapes	Balazs Varga	3	-1/+20
12 days	tdf#101322 Chart OOXML Export: fix missing subtitle	Balazs Varga	5	-10/+60
2019-08-14	tdf#124243 fix import of deleted X axis of 3D charts	Balazs Varga	3	-3/+20
2019-08-07	tdf#125337 Chart DOCX Import: fix missing empty data series columns	Balazs Varga	3	-1/+29
2019-08-05	tdf#126686 Chart: fix crash	Balazs Varga	1	-1/+1
2019-08-01	tdf#126244 Chart view: fix rotation of complex category labels	Balazs Varga	3	-3/+31
2019-07-24	tdf#114166 DOCX chart import: fix missing complex categories	Balazs Varga	8	-21/+94
2019-07-15	tdf#126193 Chart OOXML: Export Multi-level category labels	Balazs Varga	6	-41/+269
2019-06-28	tdf#126115 Chart OOXML Export: fix individual marker	Balazs Varga	4	-10/+35
2019-06-26	tdf#126033 chart: hide removed marker of a single data point	Balazs Varga	3	-1/+20
2019-06-21	tdf#124817 OOXML chart import: fix missing symbols	Balazs Varga	3	-1/+41
2019-06-18	tdf#111824 Chart OOXML Export: fix Z axis labels	Balazs Varga	3	-0/+16
2019-06-12	tdf#75659 Chart: fix empty legend at not available legend names	Balazs Varga	3	-2/+36
2019-06-04	Related: tdf#48041 Chart: vertical value axis labels	Balazs Varga	1	-22/+21
2019-06-03	tdf#48041 Chart: do not duplicate major value	Balazs Varga	3	-0/+87
2019-05-22	tdf#125335 fix order of bar chart legend names in top-bottom positions	Balazs Varga	3	-2/+25
2019-05-18	tdf#125334 Chart: allow text break in bar chart axis labels	Balazs Varga	4	-9/+38
2019-05-14	tdf#100084 XLSX import: fix missing charts clean-up	Balazs Varga	2	-5/+6
2019-05-13	tdf#100084 XLSX import: fix missing charts	Balazs Varga	3	-1/+25
2019-04-26	tdf#124796 Fix chart primary Y axis scaling	Balazs Varga	3	-2/+40
2019-04-17	tdf#123828 XLSX combined chart export: fix order of axis types	Balazs Varga	4	-7/+68
2019-04-15	tdf#114181 XLSX combined chart: fix swapped primary and secondary axes etc.	Balazs Varga	3	-4/+58
2019-04-08	tdf#124466 XLSX: fix broken export by removing chart type data redundancy	Balazs Varga	3	-2/+38
2019-04-02	tdf#124463 XLSX export: fix splitDataSeriesByAxis	Balazs Varga	3	-0/+33
2019-03-18	tdf#81437 XLSX import: fix missing chart border	Balazs Varga	5	-4/+53
2019-02-20	Related: tdf#122226 OOXML Chart Import data label separator	Balazs Varga	1	-1/+5
2019-02-15	tdf#97575 Chart OOXML: Export ShapeProps of Error Bars	Balazs Varga	3	-0/+21
2019-02-13	tdf#123400 OOXML Chart: Export Data Label Separator	Balazs Varga	3	-0/+31
2019-02-12	tdf#122226 OOXML Chart Import: data label new line separator	Balazs Varga	3	-1/+21
2019-01-30	tdf#122915 Chart OOXML: Export individual data label prop	Balazs Varga	3	-2/+16
2019-01-18	tdf#122800 Chart view: Textbreak of X axis labels	Balazs Varga	3	-2/+25
2019-01-07	tdf#122091 OOXML Import: Automatically break of X Axis labels	Balazs Varga	5	-2/+48
2018-12-18	tdf#122090 Chart: Fix OOXML export of X axis labels rotation	Balazs Varga	3	-1/+25
2018-12-18	tdf#108021 Chart: Fix text break of column chart X axis	Balazs Varga	5	-1/+43
2018-12-14	Related: tdf#118705 Fix Column/Bar Chart data label placement	Balazs Varga	1	-3/+20
2018-12-07	tdf#114163 Chart: fix order of legend names	Balazs Varga	6	-4/+47
2018-11-19	Related: tdf#108104 Fix nitpick inline in ShapeProperty enum class	Balazs Varga	1	-6/+6
2018-11-18	tdf#94231 OOXML Import: Fix disappeared Hatch Background Color	Balazs Varga	5	-8/+29
2018-11-05	tdf#108104 OOXML Import: Fix Hatch fill in Charts	Balazs Varga	9	-14/+78
2018-10-09	tdf#120341 OOXML export: fix export of Pie Chart Wall Borderline	Balazs Varga	3	-0/+21
2018-09-21	tdf#118705 Fix Column and Bar Chart data label placement	Balazs Varga	1	-2/+8
2018-09-11	tdf#108028 OOXML: Fix line breaking of chart title	Balazs Varga	1	-0/+1
2018-09-11	tdf#119617 Fix export of Autofit property of shapes to PPTX	Balazs Varga	3	-0/+22
2018-08-30	tdf#119562 Fix export of AutoFit property of shapes to XLSX	Balazs Varga	3	-1/+17
2018-08-23	tdf#94502 Fix overlap export of Stacked Bar Chart to *.xlsx	Balazs Varga	3	-1/+39
2018-08-10	tdf#96469 Export Data Point properties in Charts on DOCX/XLSX	Balazs Varga	4	-0/+91

Developments at NISZ

Tips & tricks behind the scenes

László Németh (nemeth@numbertext.org)

Short bio: László Németh (45). Code contributor since 2002. Working for Collabora Productivity and for NISZ since July, 2018.

Main results: Hunspell spell checker (<https://github.com/hunspell/hunspell/>),

Lightproof grammar checker (

<https://cgit.freedesktop.org/libreoffice/lightproof>), LibreLogo (

<http://librelogo.org/en/>), stemming and morphological generation in

thesauri (https://bz.apache.org/ooo/show_bug.cgi?id=19563), special

hyphenation algorithm and library (

<https://www.tug.org/TUGboat/tb27-1/tb86nemeth.pdf>),

Numbertext/Moneytext (<http://www.numbertext.org>) & “spell out” chapter numbering (

https://wiki.documentfoundation.org/ReleaseNotes/6.1#New_.E2.80.9Cspell_out.E2.80.9D_chapter_numbering_styles

) and NatNum 12 (

https://wiki.documentfoundation.org/ReleaseNotes/6.1#New_.E2.80.9Cspell_out.E2.80.9D_NatNum_modifier

), custom spelling dictionary with affixation (

https://wiki.documentfoundation.org/ReleaseNotes/6.0#.E2.80.9CGrammar_By.E2.80.9D_spell_checking

), custom hyphenation patterns with special hyphenation (

https://wiki.documentfoundation.org/ReleaseNotes/4.3#Non-standard_hyphenation_support_in_user-defined_dictionaries

), extension of different dictionaries for Hungarian and English.

Other interesting results: improved typography in LibreOffice by new

Graphite fonts (<http://numbertext.org/linux/>); fixing top-wanted issues, like wholly colored pages (

https://bugs.documentfoundation.org/show_bug.cgi?id=33041) and

copying between Calc and Writer (

https://bugs.documentfoundation.org/show_bug.cgi?id=37223).

~170 bug fixes

- From July 2018 to September 2019
 - ~100 bug fixes by me (OOXML, track changes)
 - ~50 bug fixes by Balázs Varga (charts)
 - ~20 bug fixes by Tünde Tóth, Gábor Kelemen, Ádám Kovács, József Szakács, Szabolcs Tóth, Bálint Sass and László Zoltán Kovács
- Recent team: 4 members + 3 newcomers + a student trainee

Short introduction of LibreOffice development at NISZ. Our mission is to fix the OOXML interoperability and LO usability bugs reported and detected in our migration projects.

Last year introduction: Building a LibreOffice development team, Tirana, 2018

http://numbertext.org/libreoffice/BuildALibreOfficeTeam_2018.pdf

Interesting for newcomer developers:

LibreOffice language technology – News & Best practices by László Németh, Tirana, 2018,

<https://github.com/hunspell/hunspell/releases/download/v1.7.0/LibreOfficeLanguageTechnologyTirana2018.pdf>

LibreOffice & eTwinning – LibreLogo in schools and teacher training programs,

http://numbertext.org/logo/LibreLogo_eTwinning_2018.pdf

How to develop LibreOffice?

- Understand the problem
- Choose the right tool
- Read the code actively
- Fix the bug or find a workaround

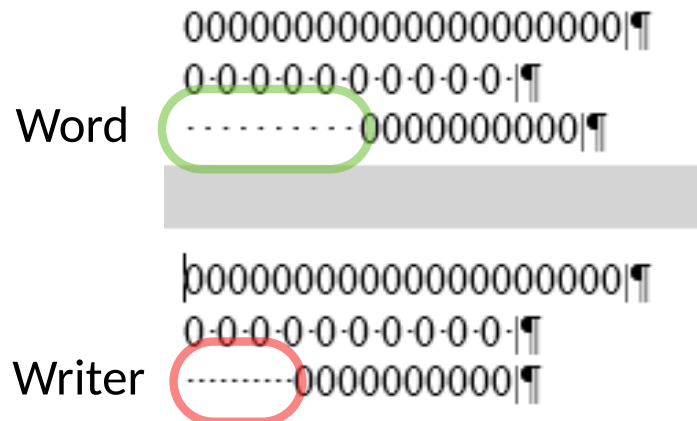
LibreOffice is one of the biggest free software projects. Thanks to the crazy bussiness idea of the 90's, its code base contains five or more different applications: a word processor, a spreadsheet and a presentation program, a graphic, a formula, an XML form and an HTML editor, a database management software, moreover, a lot of filters, including a PDF generator, the UNO programming interface, embedded Basic and Python macro languages etc.

Nobody knows the whole software, so the secret of the successful LibreOffice development is the **effective self-study of the problem and the related code base.**

Our developers have no the magic (average) 10 000 hours mastering LibreOffice development, so we have to focus on learning only the crucial part of the problem. This is the main topic: to avoid the future pitfalls...

Understand the problem

- In RTF documents, “size of spaces is different in Word and Writer”
- UI TESTING: only in space sequences!
- +1/6 em



The first typical pitfall is misunderstanding the problem. **Collect every possible questions about the problem, and answer it in LibreOffice AND in MSO.** It's much cheaper to double the time of the problem analysis 1) in the program interfaces and 2) in the test files, than making pointless or inefficient code reading and development.

In this example, only with text editing, we were able to detect the real problem, measuring also the exact difference between the lengths of the RTF space sequences printed in MSO and LO:

The original bug report was ambiguous: “size of spaces is different in Word and Writer”. Using the same amount of spaces and zeros with different combinations it became unambiguous: only consecutive spaces get extra length in MSO.

Trying to add different length Unicode space characters in the editor, the extra length was exactly SIX-EM-SPACE per space character (see Unicode input methods and Unicode spaces).

Choose the right tool

- LO/MSO/Google Docs etc.:
 - UI (editing, settings, options)
 - Import tests
 - Export tests
- XML (OOXML or ODF) analysis
 - Create test pairs with a single difference
 - Compare them by odiff or Open XML SDK

The best tool for analysing an MSO interoperability problem is MSO itself and OOXML files generated by MSO. For example, using only MSO, 1) Create an OOXML base line and 2) make only a single editing difference reproducing the problem or changing the related OOXML part 3) save it as a second test file. 4) Compare them using Open XML SDK on Windows or odiff on Linux.

Logically, to analyze export problems of LibreOffice, you must check the OOXML files exported by LibreOffice. But the first thing is to find the related OOXML elements or attributes using only MSO and an XML comparison tool.

<http://numbertext.org/odiff>

- odiff file1.docx file2.docx
- odiff file1.odt file2.odt
- odiff file1.odt # show XML content

```
<w:r>
  <w:rPr/>
  <w:t>Lorem ipsum dolor sit amet, consectetur adipiscing elit. elit.</w:t>
</w:r>
<w:r>
  <w:rPr/>
  <w:commentReference w:id="0"/>
</w:r>
<w:r>
  <w:rPr/>
  <w:t xml:space="preserve"> Vestibulum consequat mi quis pretium semper. Proin
luctus orci ac neque venenatis, quis commodo dolor posuere. Curabitur dignissim sapien
quis cursus egestas. Donec blandit auctor arcu, nec pellentesque eros molestie eget.
In consectetur aliquam hendrerit. Sed cursus mauris vitae ligula pellentesque, non pel
lentesque urna aliquet. Fusce placerat mauris enim, nec rutrum purus semper vel. Praes
```

Odiff is a simple comparison tool to display the difference of all xml and refs files of OOXML (and ODF) documents.

This part of the odiff output shows a change in the document.xml of the DOCX file: we added a comment to a paragraph, and the file format stores it (partially) as a w:commentReference element with the ID of (the text of) the comment.

Create your own tools: odiff is also an example for creating our own tools to speed up our learning process and development. (The other own notable command line tool is “note”, a logging and archiving tool for code pointers, parts and gdb backtraces.)

Check not only the OOXML difference, but read the OOXML standard, too.

(To solve the problem, you have to check the OpenDocument standard, too and its LibreOffice implementation.)

Read the code *actively*

- git grep commentReference
 sw/source/filter/ww8/docxattributeoutput.cxx:
 m_pSerializer->singleElementNS(XML_w,
 XML_commentReference, FSNS(XML_w, XML_id
), OString::number(nId));
- SAL_DEBUG("Comment ID: " << nId);
- soffice --convert-to docx comment.odt
 debug:16080:16080: Comment ID: 0

Forget <http://opengrok.libreoffice.org> and <http://api.libreoffice.org>, and start to use **git grep**, check the header files for the suspicious member variables and methods. Use **git log** to find 1) the place of the related code parts, 2) over-ready commits for code reuse and 3) for understanding the code itself: it's extremely useful not only for fixing regressions.

Use SAL_DEBUG frequently to check 1) we are in the right place during code reading and problem solving 2) check quickly the code execution.

Use a debugger to read the code by creating backtraces and step-by-step code execution (see TDF wiki page about debugging).

Try to figure out how the related code parts work and check your guess.

Command line execution of LibreOffice is faster.

Fix the bug

- Code reuse
 - git grep, git log
- Keep it stupid simple (KISS)
- Break the problem to a few parts
 - https://en.wikipedia.org/wiki/Pareto_principle#In_computing
- Create unit tests (copy and modify)

If we are lucky, the solution for our problem has already been in the git repository. Mastering git grep and git log is one the fastest ways to fix a problem.

We have multiple ways to fix a bug. Finding the best is sometimes the hardest, and usually we haven't got 3 man-months to fix a bug. Our aim is to find a solution that is good enough. The Pareto principle Wikipedia page shows that the big software companies do the same.

Delegate the problem using gerrit: ask your mentor and the experts to review your initial commit. They could help to create the best solution for the problem.

To accept your commit: add a 1) working unit test to it and a 2) compact commit description with a 3) straightforward 1-line summary, possibly with the magic word "fix".

Unit tests (sw/qa, sc/qa...)

- XPath (document format)
- Inner API
- UNO API
- GDIMetaFile (page layout as an image)
- UI (editing, ie. uiwriter)
- UI dialog windows/Python

The simplest thing is searching (using git grep or test file reading), copying and modifying an existing unit test.

There are a lot of types of unit tests. For OOXML interoperability bugs, we can extend the ooxmlexport unit tests files under sw/qa and sc/qa. The most frequently used for this kind of problems is the XPath testing. The biggest problems could be the bad XPath path here. To fix a bad path: 0) remove the end of the path to get a working Xpath assertion 1) don't leave the character / at the end of the path. 2) and try to extend it element by element carefully, choosing the correct bracket selection [nth], when the selected level contains more elements.

Speed up development

- Growth mindset
- Continuous mentoring
- Use gerrit and (libreoffice-dev or company level) developer channels

Growth mindset is the working attitude of the problem solver: knowing that we can solve the task, because we can learn everything about the problem, and its (unknown) solution. With growth mindset, even the most complex task could be an interesting and entertaining puzzle, without it, even the simplest one could be a nightmare.

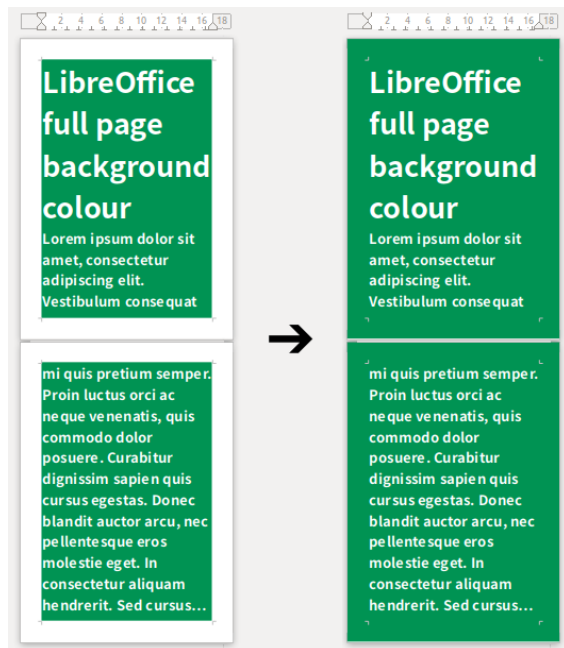
See

https://www.ted.com/talks/carol_dweck_the_power_of_believing_that_you_can_improve

One of the most important task of mentoring is to keep and strengthen the growth mindset of the mentored colleagues

Example

- Colored pages in Writer



Transferring page layout of LibreOffice Writer to the 21st century: allowing whole page background color. Using git grep, git log and a few SAL_DEBUGs, I was able to narrow the problem and its solution only to a single file in the code base. MSO do the same since decades (but only for all pages, so Writer is better now, because we can use different background colors, gradients and tiled bitmaps for pages with different page styles).

Solution

```
diff --git a/sw/source/core/layout/paintfrm.cxx b/sw/source/core/layout/paintfrm.cxx
index 1adf941..e920046 100644
--- a/sw/source/core/layout/paintfrm.cxx
+++ b/sw/source/core/layout/paintfrm.cxx
@@ -6179,7 +6179,15 @@ void SwFrame::PaintSwFrameBackground( const SwRect &rRect, const SwPageFrame *pP
    }
    else
    {
-        ::lcl_CalcBorderRect( aRect, this, rAttrs, false, gProp);
+        if ( bPageFrame )
+        {
+            aRect = getFrameArea();
+        }
+        else
+        {
+            ::lcl_CalcBorderRect( aRect, this, rAttrs, false, gProp);
+        }

        if ( (IsTextFrame() || IsTabFrame()) && GetPrev() )
        {
            if ( GetPrev()->GetAttrSet()->GetBackground() == GetAttrSet()->GetBackground() &&
```

The whole fix consists of only a few program lines. Fixing the problem in layout/core resulted also working PDF export, allowing creating up-to-date digital media using LibreOffice Writer.

The OpenDocument standard doesn't cover the area of the background of page styles of the text documents.

There was a proposed standard by Regina Henschel to extend the ODF to keep the (ugly, non typographical standard) white margins of the old ODF documents.

In my opinion, old layout was a bad and unused practice, moreover, keeping it is bad for fixing a more important problem (more resources weren't allowed for this particular problem at NISZ).

So the proposed few-line fix was merged only with the development branch to take time to discuss the possible standardization, leaving the time to solve the problem. It solved it in LO 6.3.

Example

Copy filtered Calc data into Writer tables

	A	B
1	name	sex
2	Aaliyah	female
5	Abby	female
9	Abigail	female
14	Ada	female
15	Adalee	female
16	Adaline	female

→

Before 6.2

Aaliyah	female
Abby	female
Abigail	female
Ada	female

6.2

Aaliyah	female
Aarav	male
Aaron	male
Abby	female
Abdell	male
Abdullah	male
Abel	male
Abigail	female
Abraham	male
Abram	male
Ace	male
Achilles	male
Ada	female

6.3

Aaliyah	female
Abby	female
Abigail	female
Ada	female

One of the most wanted bug fix (also in bugzilla of [OpenOffice.org/LibreOffice](https://openoffice.org/LibreOffice) and in migration projects) was to support native copy from Calc to Writer tables. Instead of the ugly and uncomfortable OLE objects, now we can extend Writer tables easily with cell data copied from Calc using Ctrl-C, Ctrl-V.

The code base of Calc and Writer are so different, that the originally planned fix by StarDivision, the original creator of the code base of StarOffice/OpenOffice.org would have taken a lot of work, leaving the problem unresolved for decades.

Solution

```

@@ -1187,6 +1192,26 @@ bool SwTransferable::Paste(SwWrtShell& rSh, TransferableDataHelper& rData, RndSt
    }
}

OfficeWriter
+
+ // tdf#37223 insert OLE table in text tables as a native text table
+ // (first as an RTF nested table, and cut and paste that to get a
+ // native table insertion, removing also the temporary nested table)
+ // TODO set a working view lock to avoid of showing the temporary nested table for a moment
+ if (bInsertOleTableInTable && EXCHG_OUT_ACTION_INSERT_STRING == nAction)
+ {
+     bool bPasted = SwTransferable::PasteData( rData, rSh, nAction, nActionFlags, nFormat,
+                                               nDestination, false, false, nullptr, 0, false, nAnchorType );
+     if (bPasted && rSh.DoesUndo())
+     {
+         SfxDispatcher* pDispatch = rSh.GetView().GetViewFrame()->GetDispatcher();
+         pDispatch->Execute(FN_PREV_TABLE, SfxCallMode::SYNCHRON);
+         pDispatch->Execute(FN_TABLE_SELECT_ALL, SfxCallMode::SYNCHRON);
+         pDispatch->Execute(SID_COPY, SfxCallMode::SYNCHRON);
+         pDispatch->Execute(SID_UNDO, SfxCallMode::SYNCHRON);
+         pDispatch->Execute(SID_PASTE, SfxCallMode::SYNCHRON);
+     }
+     return bPasted;
+ }
+
+

```

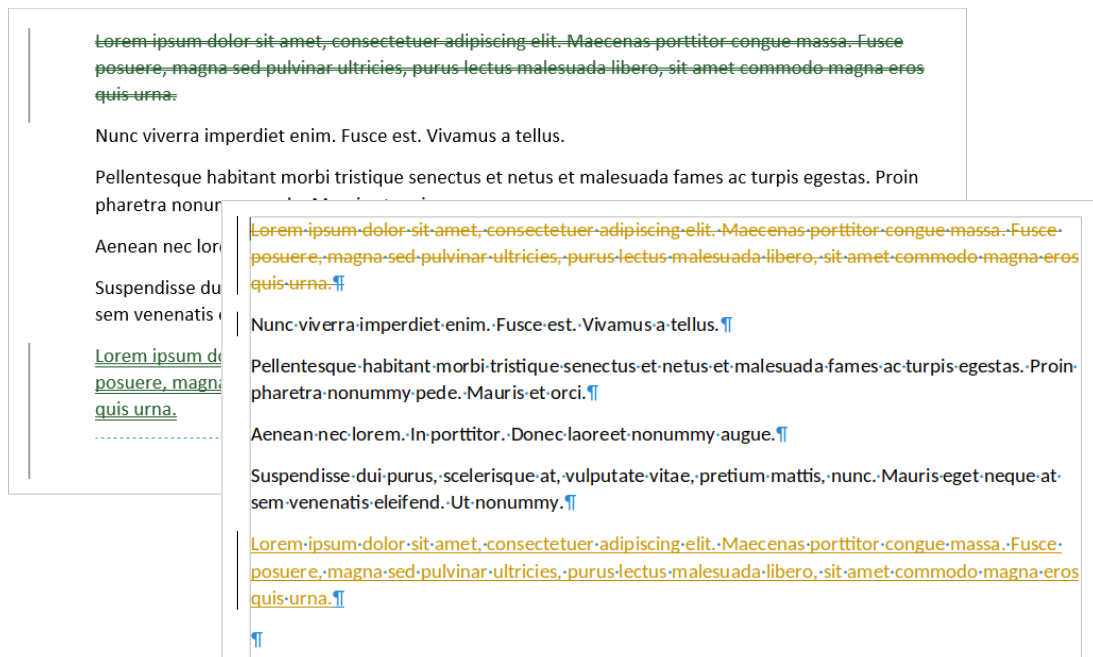
Instead of developing a new filter to handle the special data transfer, I've found a 2-step workaround significantly simplifying the solution. The fix inserts the clipboard data as a temporary table at the place of the text cursor, and copies back its content as native text table data to the clipboard, remove the table, and inserts the native text table data in the native text table.

The slide shows the main part of the original fix.

(Note: there is a TODO yet about the visible temporary table, but the non-debug build of LibreOffice doesn't show the problem.)

With recent rewrite of the commit in LibreOffice 6.3, this works well with copying the result of Calc filtering, ie. inserting only the visible cells of the selection.

Example



One of our most complex problems, to add a new feature to LibreOffice, supporting text moving during change tracking. Double green underline shows it in Word. Breaking the problem resulted the following TODO list:

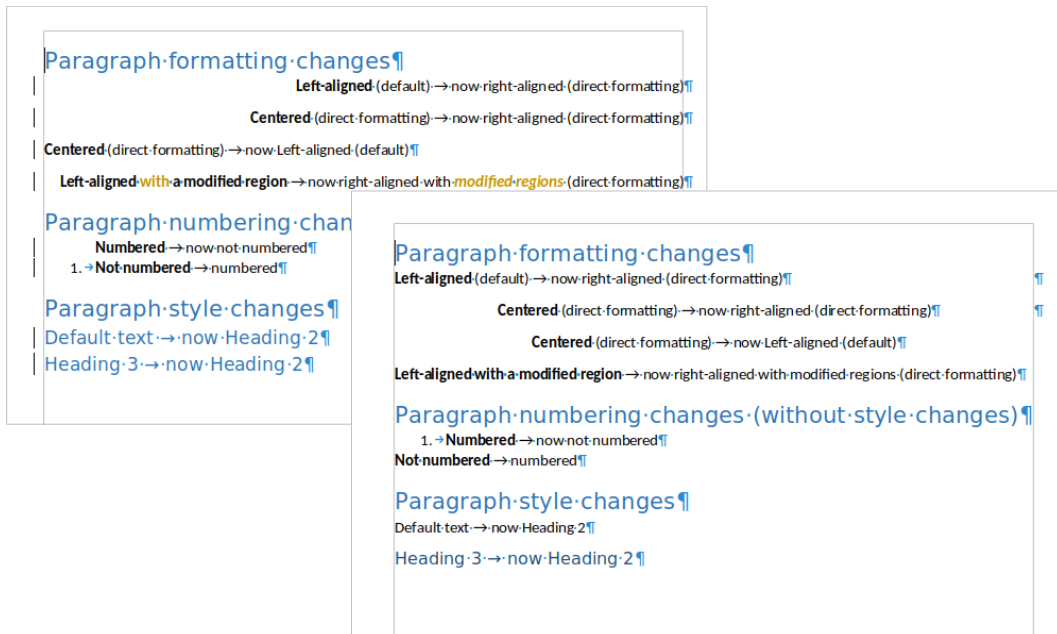
- 1) Keep the actual content of the document (LibreOffice imported also the removed text parts as part of the actual text content).
- 2) Use text deletion/insertion temporarily to show and handle text moving.
- 3) Show text moving in the user interface of LibreOffice, too.

Solution

```
diff --git a/writerfilter/source/dmapper/DomainMapper_Impl.cxx
index 8a9c038..321866d 100644
--- a/writerfilter/source/dmapper/DomainMapper_Impl.cxx
+++ b/writerfilter/source/dmapper/DomainMapper_Impl.cxx
@@ -1860,9 +1860,11 @@ void DomainMapper_Impl::CreateRedline(un
     case XML_mod:
         sType = getPropertyNames( PROP_FORMAT );
         break;
+    case XML_moveTo:
+    case XML_ins:
         sType = getPropertyNames( PROP_INSERT );
         break;
+    case XML_moveFrom:
+    case XML_del:
         sType = getPropertyNames( PROP_DELETE );
         break;
```

We are lucky sometimes. Developing part 1 (fix broken document content) and part 2 (import text moving from/to as deletion/insertion) only consisted of the following two lines.

Example



We fixed most of the reported DOCX change tracking bugs, also other interoperability and usability problems of this area, for example, handling of DOCX text formatting (see screenshot).

Our aim to create a stable tool for top of change tracking: editing legal texts.

Recent improvements of Michael Stahl (CIB) opened the way to achieve this, so we started to fix the most serious problems and corner cases of change tracking. Now LibreOffice keeps formatting change data during DOCX import/export, it can use these to reject character and paragraph formatting changes, including changes of paragraph styles and numbering.

12 hours	tdf#124491 DOCX: import tracked changes of empty paragraphs	László Németh	6	-5/+35
7 days	tdf#105485 DOCX: import deleted comments as deleted	László Németh	3	-1/+10
9 days	tdf#123703 strip six-em-space (U+2006) at line break	László Németh	4	-7/+9
9 days	tdf#123703 RTF import: fix length of space character sequence	László Németh	4	-1/+68
10 days	tdf#127101 Change tracking: fix format reject at paragraph join	László Németh	4	-6/+33
10 days	tdf#127101 Change tracking: reject format at paragraph join	László Németh	2	-17/+128
11 days	tdf#127097 Change tracking: change style of whole deletion at paragraph join	László Németh	2	-7/+46
2019-08-02	tdf#123702 RTF/DOCX default 1" left/right page margins	László Németh	2	-7/+7
2019-08-02	tdf#126243 DOCX change tracking: pStyle needs simplified style id	László Németh	4	-21/+28
2019-08-01	tdf#118699 change tracking: fix accept before empty line	László Németh	2	-53/+57
2019-08-01	tdf#118699 change tracking: don't number empty lines	László Németh	1	-1/+1
2019-08-01	tdf#125990 change tracking: remove text join workaround	László Németh	5	-92/+26
2019-07-23	tdf#126516 DOCX change tracking: fix paragraph style rejection	László Németh	4	-13/+44
2019-07-19	tdf#125546 DOCX import: fix overgrowth of change tracking entries	László Németh	5	-14/+39
2019-07-18	tdf#125894: DOCX: import tracked changes in frames	László Németh	5	-3/+59
2019-07-18	Revert "tdf#123627 DOCX import: fix relative hyperlinks to documents"	László Németh	3	-7/+6
2019-07-12	tdf#126245 DOCX export: fix losing of tracked numbering changes	László Németh	6	-1/+22
2019-07-12	tdf#126245 DOCX: reject tracked paragraph numbering changes	László Németh	3	-4/+43
2019-07-10	tdf#121663 DOCX import: fix missing space in line numbering	László Németh	4	-2/+16
2019-07-09	tdf#126243 DOCX: export/reject tracked paragraph style change	László Németh	7	-28/+59
2019-07-06	tdf#120338 DOCX: reject other paragraph formatting changes	László Németh	1	-6/+8
2019-07-05	tdf#120338 DOCX: add rejection of paragraph formatting changes	László Németh	3	-0/+60
2019-07-03	tdf#126206 DOCX: add rejection of character formatting changes	László Németh	4	-8/+48
2019-06-24	tdf#119571 fix style & numbering at tracked deletion	László Németh	2	-53/+131
2019-06-24	tdf#54819 keep style & numbering at tracked deletion	László Németh	4	-1/+129
2019-06-17	tdf#125916 DOCX import: fix outline numbering at tracked deletion	László Németh	3	-0/+19
2019-06-14	tdf#125881 DOCX import: handle list level after tracked deletion	László Németh	3	-4/+23
2019-06-13	tdf#89991 DOCX: import Show changes from older formats	László Németh	1	-2/+10
2019-06-12	tdf#89991 DOCX: import/export Show changes mode	László Németh	8	-5/+48
2019-06-07	sanitize LibreLogo calls	László Németh	1	-1/+50
2019-06-06	tdf#115212 DOCX export: fix ToC with tracked deletion	László Németh	3	-0/+9
2019-06-04	tdf#70234 DOCX export: fix only expanded fields	László Németh	1	-2/+2
2019-06-04	tdf#70234 DOCX: export tracked deletion of fields	László Németh	3	-6/+32
2019-06-04	tdf#70234 DOCX: import tracked deletion of fields	László Németh	3	-0/+8
2019-06-03	tdf#120412 char formatting UI: clean-up DFLT_ESC_AUTO	László Németh	4	-9/+21
2019-06-03	tdf#112567 XLSX export: correct built-in names	László Németh	3	-1/+55
2019-05-29	tdf#120412 character formatting UI: allow >100% raising	László Németh	1	-0/+3
2019-05-28	tdf#92335 DOCX: fix multiplying of "ListLabel" styles	László Németh	3	-0/+14
2019-05-28	DOCX import: clean-up FinalizeImport() of change tracking	László Németh	4	-4/+24
2019-05-27	tdf#115557 DOCX import: don't multiply charts in footnotes	László Németh	3	-9/+23
2019-05-24	tdf#125319 DOCX track changes: don't change numbering	László Németh	1	-3/+10
2019-05-22	tdf#121784 DOCX import: fix change tracking of footnotes	László Németh	3	-1/+21
2019-05-21	tdf#125310 change tracking: don't copy page break	László Németh	3	-1/+58
2019-05-20	tdf#125310 change tracking: don't copy page break	László Németh	3	-1/+64
2019-05-15	tdf#118699 DOCX import: don't add numbering	László Németh	4	-0/+56
2019-05-11	tdf#125187 DOCX track changes: fix w:del within w:ins	László Németh	2	-5/+5
2019-05-08	tdf#116501 fix freezing at embedded text tables	László Németh	6	-2/+44
2019-05-02	tdf#112567 XLSX export: fix broken built-in names	László Németh	3	-1/+37
2019-04-29	tdf#123460 DOCX track changes: moveFrom completely	László Németh	5	-8/+47
2019-04-26	tdf#67207 DOCX mail merge: fix export/import of database fields	László Németh	8	-21/+120

We fixed the serious text content changes during DOCX import and most of the WYSWYG anomalies of the Show changes mode.

3 days	tdf#103988 OOXML export: fix bubble chart MSO interoperability	Balazs Varga	2	-3/+7
3 days	tdf#114657 OOXML chart export: fix broken chart with NaN X value	Balazs Varga	3	-18/+22
3 days	tdf#124083 OOXML import: fix broken charts with missing value X	Balazs Varga	3	-2/+27
11 days	tdf#127030 PPTX export: keep fixed size of legacy shapes	Balazs Varga	3	-1/+20
12 days	tdf#101322 Chart OOXML Export: fix missing subtitle	Balazs Varga	5	-10/+60
2019-08-14	tdf#124243 fix import of deleted X axis of 3D charts	Balazs Varga	3	-3/+20
2019-08-07	tdf#125337 Chart DOCX Import: fix missing empty data series columns	Balazs Varga	3	-1/+29
2019-08-05	tdf#126686 Chart: fix crash	Balazs Varga	1	-1/+1
2019-08-01	tdf#126244 Chart view: fix rotation of complex category labels	Balazs Varga	3	-3/+31
2019-07-24	tdf#114166 DOCX chart import: fix missing complex categories	Balazs Varga	8	-21/+94
2019-07-15	tdf#126193 Chart OOXML: Export Multi-level category labels	Balazs Varga	6	-41/+269
2019-06-28	tdf#126115 Chart OOXML Export: fix individual marker	Balazs Varga	4	-10/+35
2019-06-26	tdf#126033 chart: hide removed marker of a single data point	Balazs Varga	3	-1/+20
2019-06-21	tdf#124817 OOXML chart import: fix missing symbols	Balazs Varga	3	-1/+41
2019-06-18	tdf#111824 Chart OOXML Export: fix Z axis labels	Balazs Varga	3	-0/+16
2019-06-12	tdf#75659 Chart: fix empty legend at not available legend names	Balazs Varga	3	-2/+36
2019-06-04	Related: tdf#48041 Chart: vertical value axis labels	Balazs Varga	1	-22/+21
2019-06-03	tdf#48041 Chart: do not duplicate major value	Balazs Varga	3	-0/+87
2019-05-22	tdf#125335 fix order of bar chart legend names in top-bottom positions	Balazs Varga	3	-2/+25
2019-05-18	tdf#125334 Chart: allow text break in bar chart axis labels	Balazs Varga	4	-9/+38
2019-05-14	tdf#100084 XLSX import: fix missing charts clean-up	Balazs Varga	2	-5/+6
2019-05-13	tdf#100084 XLSX import: fix missing charts	Balazs Varga	3	-1/+25
2019-04-26	tdf#124796 Fix chart primary Y axis scaling	Balazs Varga	3	-2/+40
2019-04-17	tdf#123828 XLSX combined chart export: fix order of axis types	Balazs Varga	4	-7/+68
2019-04-15	tdf#114181 XLSX combined chart: fix swapped primary and secondary axes etc.	Balazs Varga	3	-4/+58
2019-04-08	tdf#124466 XLSX: fix broken export by removing chart type data redundancy	Balazs Varga	3	-2/+38
2019-04-02	tdf#124463 XLSX export: fix splitDataSeriesByAxis	Balazs Varga	3	-0/+33
2019-03-18	tdf#81437 XLSX import: fix missing chart border	Balazs Varga	5	-4/+53
2019-02-20	Related: tdf#122226 OOXML Chart Import data label separator	Balazs Varga	1	-1/+5
2019-02-15	tdf#97575 Chart OOXML: Export ShapeProps of Error Bars	Balazs Varga	3	-0/+21
2019-02-13	tdf#123400 OOXML Chart: Export Data Label Separator	Balazs Varga	3	-0/+31
2019-02-12	tdf#122226 OOXML Chart Import: data label new line separator	Balazs Varga	3	-1/+21
2019-01-30	tdf#122915 Chart OOXML: Export individual data label prop	Balazs Varga	3	-2/+16
2019-01-18	tdf#122800 Chart view: Textbreak of X axis labels	Balazs Varga	3	-2/+25
2019-01-07	tdf#122091 OOXML Import: Automatically break of X Axis labels	Balazs Varga	5	-2/+48
2018-12-18	tdf#122090 Chart: Fix OOXML export of X axis labels rotation	Balazs Varga	3	-1/+25
2018-12-18	tdf#108021 Chart: Fix text break of column chart X axis	Balazs Varga	5	-1/+43
2018-12-14	Related: tdf#118705 Fix Column/Bar Chart data label placement	Balazs Varga	1	-3/+20
2018-12-07	tdf#114163 Chart: fix order of legend names	Balazs Varga	6	-4/+47
2018-11-19	Related: tdf#108104 Fix nitpick inline in ShapeProperty enum class	Balazs Varga	1	-6/+6
2018-11-18	tdf#94231 OOXML Import: Fix disappeared Hatch Background Color	Balazs Varga	5	-8/+29
2018-11-05	tdf#108104 OOXML Import: Fix Hatch fill in Charts	Balazs Varga	9	-14/+78
2018-10-09	tdf#120341 OOXML export: fix export of Pie Chart Wall Borderline	Balazs Varga	3	-0/+21
2018-09-21	tdf#118705 Fix Column and Bar Chart data label placement	Balazs Varga	1	-2/+8
2018-09-11	tdf#108028 OOXML: Fix line breaking of chart title	Balazs Varga	1	-0/+1
2018-09-11	tdf#119617 Fix export of Autofit property of shapes to PPTX	Balazs Varga	3	-0/+22
2018-08-30	tdf#119562 Fix export of AutoFit property of shapes to XLSX	Balazs Varga	3	-1/+17
2018-08-23	tdf#94502 Fix overlap export of Stacked Bar Chart to *.xlsx	Balazs Varga	3	-1/+39
2018-08-10	tdf#96469 Export Data Point properties in Charts on DOCX/XLSX	Balazs Varga	4	-0/+91

We fixed a lot of OOXML chart interoperability issues. With help of my colleagues, the LibreOffice developer and user community, we are on the right track to fix more complex problems, too.

Thanks for your help!