

GSoC project: Improving RTF Export Presentation of a GoOO student

Miklos Vajna

22 October 2010

- Background
- Activities in GoOO before SoC
- Project this summer: development of a new RTF export filter
 - developer side
 - user side

- I'm a student from Budapest University of Technology and Economics, Hungary
- A few project I am interested in:
 - Frugalware Linux - a distribution
 - BitlBee - an IM to IRC gateway (Skype module)
 - git - I wrote the current git-merge command
 - swig - the binding generator (PHP director support)
 - LibreOffice - packaging, RTF export filter

- Packager for Frugalware Linux
- Minor build system fixes
 - trivial support for newer gcj versions
 - git-related patches
- No C++ coding (steep learning curve)

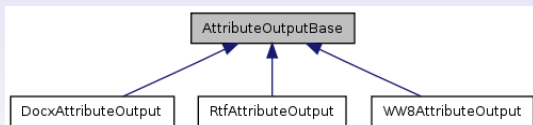
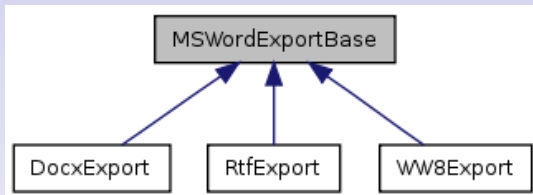
RTF Export Development

Summary

- idea: the concept of RTF is very similar to doc/docx (Microsoft invented them), just with a different markup
- Novell already created `MSWordExportBase`
- target: to support everything which was provided by the old filter, smaller size, new features

RTF Export Development

The common base



- `MSWordExportBase`: tries to map Writer concepts to MSO
- `AttributeOutputBase`: 120+ methods for different resources

- RtfExportFilter: glue between RtfExport and UNO
- RtfImportFilter: glue between the old RTF import and UNO
- RtfSdrExport: handles drawings
- RtfFilter in writerfilter: calls RtfExportFilter and RtfImportFilter via UNO

RTF Export Development

Is the goal reached?

- No regressions compared to the old filter: mostly
 - still needs more testing, but it's enabled by default
- Smaller code:
 - 47 files changed
 - 7567 insertions(+), 6981 deletions(-)
 - More code - due to better structured code, new features.
- More features: far from lossless conversion, but a number of new features

- German comments
- No test files for the old filter
- Can't wait for the moment when split build will be recommended for development
- Non-intuitive API's

- To send over an SvStream through UNO:

`utl::OStreamWrapper()` to wrap it,

`utl::UcbStreamHelper::CreateStream()` to unwrap

- No common base for header / footer - duplicated `IsActive()` method:

```
class SW_DLLPUBLIC SwFmtHeader: public
    SfxPoolItem, public SwClient {...}
```

```
class SW_DLLPUBLIC SwFmtFooter: public
    SfxPoolItem, public SwClient {...}
```

- Getting the streams from a media descriptor:

- `MediaDescriptor::PROP_STREAMFOROUTPUT()` - output
- `MediaDescriptor::PROP_INPUTSTREAM()` - input

- `fromhex.py` - reads a hexdump from RTF and writes it as a binary
- `prettyprint.py` - pretty-prints an RTF file
- `oodocdiff.sh` - to test
 - sadly it's mostly useless due to character kerning and other changes

- Cloned ooo-test-files from Cedric / Kohei
- Added 25 test ODT files
- Tested with headless writer + jodconverter using UNO
 - that can be now replaced with the batch conversion patches

- Chicken and egg problem: pointless to read specs from start to end, but how to just use it as a reference?
- Necessary resources:
 - Rich Text Format (RTF) Specification, version 1.9.1 (Word 2007)
 - Object Linking and Embedding (OLE) Data Structures
 - ISO/IEC 29500-1:2008 - OOXML spec
 - Word Binary File Format (.doc) Structure Specification

Before:

```
Hello world! 1¶  
reference to footnote: Error: Reference source not found¶
```

After:

```
Hello world! 1¶  
reference to footnote: 1¶
```

RTF Export Development

Character properties

Before:

```
expanded·space·by·5·and·10pt¶  
VA·no·kerning:·VA¶
```

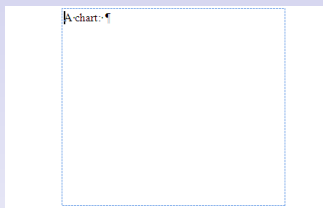
After:

```
expanded·s p a c e ·by·5·and·10pt¶  
VA·no·kerning:·VA¶
```

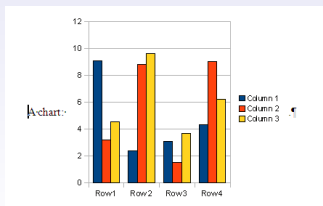
RTF Export Development

Charts (and other OLE objects)

Before:



After:













RTF Export Development

Drawings

Before:

```
Rectangle.¶  
¶  
Ellipse.¶  
¶  
¶  
Freeform-line.¶  
¶  
Line.¶  
¶  
Basic-shape-examples.¶  
  → regular-pentagon, cross, cube¶  
¶  
Text.¶  
¶  
Vertical-text.¶  
¶  
Magnetic-disc-with-formatted-text.¶
```

After:

```
Rectangle.¶   
¶  
Ellipse.¶   
¶  
¶  
Freeform-line.¶   
¶  
Line.¶   
¶  
Basic-shape-examples.¶     
  → regular-pentagon, cross, cube¶  
¶  
Text.¶   
¶  
Vertical-text.¶   
¶  
Magnetic-disc-with-formatted-text.¶ 
```

Before:

This one is selected:

This is a textbox:

Listbox:

After:

This one is selected:

This is a textbox:

Listbox:

Drop-Down Form Field Options

Drop-down item:

Items in drop-down list:

- First entry
- Second entry
- Third entry

Buttons: Add >>, Remove, Move (up/down arrows)

Before:

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis rhoncus imperdiet velit, sed luctus augue pellentesque a. Cras semper purus eu est porta euismod. Sed auctor laoreet erat, pretium condimentum libero condimentum sit amet. Aenean lectus purus, blandit ut volutpat eget, consequat eu leo. In dignissim rhoncus libero ut sollicitudin. Donec egestas viverra mi eu consequat. Maecenas malesuada sagittis posuere. Curabitur facilisis, dui nec cursus viverra, felis arcu vulputate magna, at fringilla est neque pretium ipsum. Duis in velit diam, bibendum vehicula diam. Cras nunc ligula, condimentum non tincidunt et, interdum sit amet enim. Sed ac felis vitae justo tempus dignissim. Praesent dui massa, imperdiet id dapibus at, egestas sed enim. Nam bibendum blandit nisi ac dignissim. Cras eget uma justo. Sed mi uma, facilisis et mollis a, rutrum ac libero. Donec pellentesque luctus posuere. ¶

After:

5 Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis rhoncus imperdiet velit, sed luctus augue pellentesque a. Cras semper purus eu est porta euismod. Sed auctor laoreet erat, pretium condimentum libero condimentum sit amet. Aenean lectus purus, blandit ut volutpat eget, consequat eu leo. In dignissim rhoncus libero ut sollicitudin. Donec egestas viverra mi eu consequat. Maecenas malesuada sagittis posuere. Curabitur facilisis, dui nec cursus viverra, felis arcu vulputate magna, at fringilla est neque pretium ipsum. Duis in velit diam, bibendum vehicula diam. Cras nunc ligula, condimentum non tincidunt et, interdum sit amet enim. Sed ac felis vitae justo tempus dignissim. Praesent dui massa, imperdiet id dapibus at, egestas sed enim. Nam bibendum blandit nisi ac dignissim. Cras eget uma justo. Sed mi uma, facilisis et mollis a, rutrum ac libero. Donec pellentesque luctus posuere. ¶

10

RTF Export Development

Mathematical Expressions

Before:

before¶

After:

before $a = \frac{b}{c}$ after¶

RTF Export Development

Editing Mathematical Expressions

before $a = \frac{b}{c}$ after

Equation

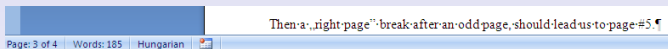
$\leq \approx$	$\Delta a b \cdot$	$\frac{\cdot}{\cdot}$	$\pm \cdot \otimes$	$\rightarrow \leftrightarrow \downarrow$	$\cdot \cdot \forall \exists$	$\notin \cap \subset$	$\partial \infty \ell$	$\lambda \omega \theta$	$\Delta \Omega \oplus$
$() []$	$\sqrt{\quad}$	$\frac{\cdot}{\cdot}$	$\Sigma \sum$	$\int \oint$	$\frac{\cdot}{\cdot}$	$\rightarrow \leftarrow$	$\dot{\cup} \dot{\cup}$	\dots	$\begin{matrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{matrix}$

RTF Export Development

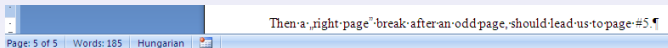
Special Page Breaks

Test: normal, right (RTF calls it 'odd'), right again (the last one will be a double break)

Before:



After:



RTF Export Development

Page Numbering

Before:

→ 1¶
[This is the first page with roman numerals.¶.....Section Break (Next Page).....

→ 2¶
This is the first page with „Default” style, with page numbering restarted.¶

After:

¶ → i¶
This is the first page with roman numerals.¶.....Section Break (Next Page).....

¶ → 1¶
This is the first page with „Default” style, with page numbering restarted.¶

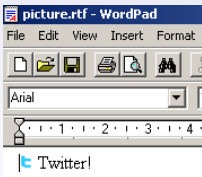
RTF Export Development

Pictures in Wordpad

Before:



After:



RTF Export Development

Post-it Field

Before:



After:



RTF Export Development

Sections: Column Break

Before:

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Column break. ¶
 Vivamus elit tellus, accumsan in congue eget, ullamcorper id odio. Ut nunc est, rutrum non facilisis ac, iaculis sit amet mauris. Curabitur ut magna ut odio vehicula scelerisque vitae at ante. Aliquam erat volutpat. Vivamus varius augue ut metus tristique id viverra nulla semper. Curabitur volutpat est tortor, sed dictum ipsum. Nam sed neque ante, nec sodales lorem. Quisque eleifend auctor pulvinar. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Duis et tempus velit. Donec porttitor nisi et est feugiat vulputate sed eget urna. Fusce facilisis, mi a suscipit venenatis, turpis lacus egestas odio, in tempus quam lacus ut est. ¶

After:

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Column break. ¶

Vivamus elit tellus, accumsan in congue eget, ullamcorper id odio. Ut nunc est, rutrum non facilisis ac, iaculis sit amet mauris. Curabitur ut magna ut odio vehicula scelerisque vitae at ante. Aliquam erat volutpat. Vivamus varius augue ut metus tristique id viverra nulla semper. Curabitur volutpat est tortor, sed dictum ipsum. Nam sed neque ante, nec sodales lorem. Quisque eleifend auctor pulvinar. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Duis et tempus velit. Donec porttitor nisi et est feugiat vulputate sed eget urna. Fusce facilisis, mi a suscipit venenatis, turpis lacus egestas odio, in tempus quam lacus ut est. ¶

RTF Export Development

Protected Sections

Before:

```
⌘.....Section Break (Continuous).....  
This section is not write-protected (without a password).⌘.....Section Break (Continuous).....  
⌘
```

After:

```
⌘.....Section Break (Continuous).....  
This section is write-protected (without a password).⌘.....End of Protected Section.....  
⌘
```

RTF Export Development

Nested Tables

Before:

Nested-table:¶		
A'□		B'□
Nested-a□	Nested-b□	□
Nested-c□	Nested-d□	
□	D'□	
A row after nested table□		F'□
¶		

After:

Nested-table:¶		
A'□		B'□
Nested-a□	Nested-b□	D'□
Nested-c□	Nested-d□	
A row after nested table□		F'□
¶		

RTF Export Development

Table of Contents

Before:

• TOC	Section Break (Continuous)	
• <u>Table of Contents</u>		
First heading	→	1
Second heading	→	1
Third heading	→	1
TOC	→	1

After:

→ TOC	Section Break (Continuous)	
• <u>Table of Contents</u>	Section Break (Continuous)	
First heading	→	1
Second heading	→	1
Third heading	→	1
TOC	→	1

Acknowledgements

In no particular order:

- Cedric and Kendy: my mentors
- Thorsten: testing ideas
- Kohei: when hacking in the night
- Bublì: when the Czech guys were not on IRC
- Petr: patching issues
- everyone else who helped on #go-oo

- **LibreOffice:**
<http://www.documentfoundation.org/develop/>
- **SoC:** <http://code.google.com/soc/>
- **New RTF export filter:** <http://cgit.freedesktop.org/~vmiklos/ooo-gsoc/>