

Smarter Ways to Track Document Changes

Attila Szűcs

Software Engineer

attila.szucs@collabora.com



Collabora
Online



Technical Day
COOL
days





The problem:

- Changes fragmented
(delete the middle of an insert)
- Multiple Reject/Accept needed
- Move is 2 change too (insert + delete)

Word Reject/Accept the whole original change at once.
(Only in simple cases)

Teinsertedxt

	Action	Author
+		first user
✓ X		second user
+		first user
+		first user



Fragmentation is needed

- Tracked Changes can become very complex (mixing changes, even non tracked.)
change parts can be separated.
- File formats store them separated also.
Should we store history? .. bad idea
(all need to be tracked / new file format / performance?)
- Our list is more clear then Word list.
(order, relation between changes)

Teinse

...

insert a whole document

...

rtedxt

	Action	Author
+		first user
∨	×	second user
	+	first user
	×	third user
∨	×	second user
	+	first user
	+	first user



Interdependent changes not handled:

Data structures support interdependent changes.

But the code does not really support it. expect that there is only 1 change in 1 place. (deleting an insertion is an exception)

- For example, the function:

```
RedlineType GetType( sal_uInt16 nPos = 0 )  
const;
```

Mostly used without parameter.

- Was not able to reject/insert an inner change.
- Format change. (exist only alone)



Many different cases:

Different cases need to be handled differently:

- Deleted insertion:
 - accept insert → Keep deletion.
 - reject insert → Deletion is meaningless.
 - reject delete → Keep insertion
- Same or different author, anonym?
- Insert line: text+EOP or EOP+text?
Move line prefer EOP+text
- move a text with insertions, deletions. (No double delete)

1. → One¶

2. → 1.5¶

1. → One¶

2. [1.] 1.5¶

Some more text to ~~test~~ move.

Nothing.



~~Some more text to test move.~~

Nothing.

Some more text to move.



Move:

Action	Author	Date
	Attila Szűcs	06/04/2024 16:19:36
	first user	06/10/2023 03:32:39
	Attila Szűcs	06/04/2024 16:19:36

- Added a new unique Move ID to deletion/insertion changes. To find their pairs/fragmented parts.
It generated in some cases, but not all. (move paragraph)
- Kept the move recognition. (To avoid regression)
(even improved it “ab”+”cd” = “a”+”bcd”)
- Implemented odt export/import
- improved docx import
(not changed export, it works a bit, but not perfect)



Format:

- Checked Format change as well, but it need a lot of work.
I did not done it.
Now, it can exist only alone, everything overwrite format changes.
- Design questions: formatting a format change.
 - Make double format?
 - Merge them to one change?
 - Overwrite the original?
- Delete, fomat-1, .. format-N, Insert
formatting a deletion, should insert format under deletion.
formatting an insertion, should insert format top of insertion.
formatting a format → top of last format



Collabora
Online



Technical Day
COOL
days



Thank you!

By Attila Szűcs



@CollaboraOffice
hello@collaboraoffice.com
www.collaboraoffice.com