

# Solid Framework 10.0.12602

## Quality Improvements

Reconstruction and API Improvements since Build 12246

Release Date - 3rd October 2021

# Main Improvements

- Modified Character Spacing Corrector

Solid Framework aims to layout words in the same way that they were in the PDF, with each line being the same, but without using hard line breaks to force this to occur. The objective is a document that looks like the PDF while still being editable.

This release includes a new version of the algorithm that is used for calculating layout.
- Removal of artifacts added during PDF creation from Excel
- NSE improvements

# Each slide has an icon that indicates how the case is tested



Tested automatically at every build



Tested manually. This may be because there is no obvious way to test, or that testing is too slow

Where there are multiple slides for a given issue, these generally represent a small subset of the many files that have been improved

# Reconstruction as Word

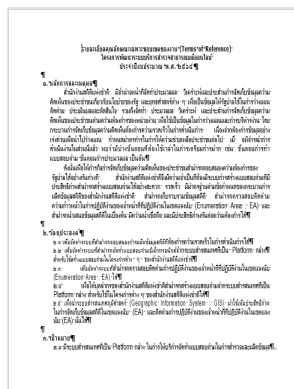
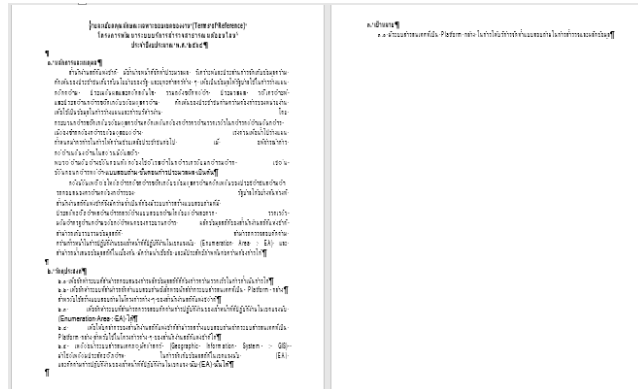
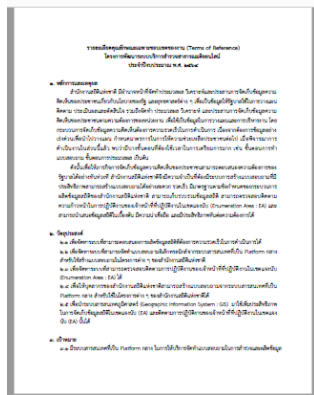


# Thai Language Combining Characters - Docx

PDF

12246

12602



Variable font properties prevent character combining - resulting in text flowing onto next page

Characters now merge, and conversion is correct

Original name	
SHA1	7ff948835ef653570c5c94561ed63ffa73daa853
FB	27793



# Arabic Language Combining Characters

PDF

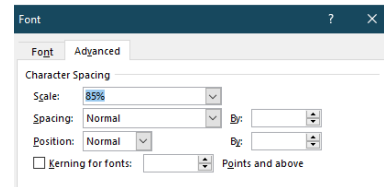
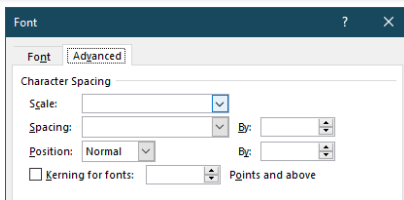
12246

12602

بين وزارة تنمية المجتمع  
شركة الامارات للتصوير والتسويق ذ.م.م  
على خدمات صيانة وتأهيل عدد 5 طابع  
بالاضافة الى 4 أجهزة تغليف حرارى.

وزارة تنمية المجتمع مع  
شركة الامارات للتصوير والتسويق ذ.م.م  
على خدمات صيانة وتأهيل عدد 5 طابع  
بالاضافة الى 4 أجهزة تغليف حرارى.

وزارة تنمية المجتمع  
شركة الامارات للتصوير والتسويق ذ.م.م  
على خدمات صيانة وتأهيل عدد 5 طابع  
بالاضافة الى 4 أجهزة تغليف حرارى.



Variable font properties  
prevent character combining

Characters now merge, and  
conversion is correct

Original name	
SHA1	272d97ddf82e35f8a0d8a3a0d64fc87ca4d27f83
FB	26877

TESTED

# Annotation Border is Now Correct

PDF



12246



Border is incorrect

12620



Conversion is correct

Original name	
SHA1	b59ee2994b72e0a2621ec48d7694b19ed2903484
FB	23420

TESTED

# Reconstruction as PowerPoint





# Leader Characters Cause Text Wrapping

PDF

SINDICATURA DE COMPTES DE CATALUNYA — INFORME 3/2000-A

2.7. CONVENIS .....	34
2.8. PRESSUPOST .....	42
2.9. CONTRACTACIÓ .....	46
3. CONCLUSIONS: OBSERVACIONS I RECOMANACIONS .....	48
4. TRÀMIT D'AL-LEGACIONS .....	51
5. ANNEXOS .....	53
5.1. BALANÇ DE SITUACIÓ A 31.12.1997 I A 31.12.1998 .....	53
5.2. COMPTA DE PÈRDUES I GUANYES. EXERCICIS 1997 I 1998 .....	55

12246

SINDICATURA DE COMPTES DE CATALUNYA — INFORME 3/2000-A

2.7. CONVENIS .....	34
2.8. PRESSUPOST .....	42
2.9. CONTRACTACIÓ .....	46
3. CONCLUSIONS: OBSERVACIONS I RECOMANACIONS .....	48
4. TRÀMIT D'AL-LEGACIONS .....	51
5. ANNEXOS .....	53
5.1. BALANÇ DE SITUACIÓ A 31.12.1997 I A 31.12.1998 .....	53
5.2. COMPTA DE PÈRDUES I GUANYES. EXERCICIS 1997 I 1998 .....	55

12620

SINDICATURA DE COMPTES DE CATALUNYA — INFORME 3/2000-A

2.7. CONVENIS .....	34
2.8. PRESSUPOST .....	42
2.9. CONTRACTACIÓ .....	46
3. CONCLUSIONS: OBSERVACIONS I RECOMANACIONS .....	48
4. TRÀMIT D'AL-LEGACIONS .....	51
5. ANNEXOS .....	53
5.1. BALANÇ DE SITUACIÓ A 31.12.1997 I A 31.12.1998 .....	53
5.2. COMPTA DE PÈRDUES I GUANYES. EXERCICIS 1997 I 1998 .....	55

The original document was created in Word using “Leaders” as part of the tab formatting. PowerPoint does not support this

Some lines wrap onto the next line

Line breaks are correct (although text layout is not quite perfect)

Original name	
SHA1	3b560400fd0c94be56ec764280313729e90b5db5
FB	21844



# Paragraph Alignment Now Correct

PDF

does not have a timeline for phasing-out, they will be reviewed periodically.

26. To enhance regional integration as well as to maintain a competitive investment area, both the Framework

First row is justified

12246

does not have a timeline for phasing-out, they will be reviewed periodically.



26.To enhance regional integration as well as to maintain a competitive investment area, both the Framework

First row is left aligned

12602

does not have a timeline for phasing-out, they will be reviewed periodically.

26.To enhance regional integration as well as to maintain a competitive investment area, both the Framework

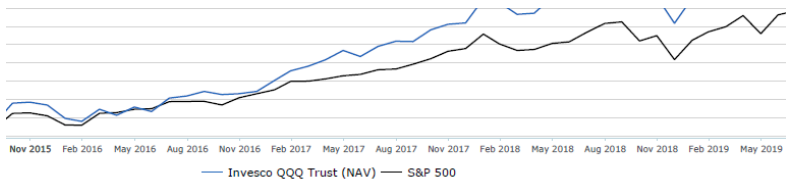
Conversion is correct

Original name	
SHA1	edfbf78a181cc1b42a5904d26a3811e7803eb274
FB	21709

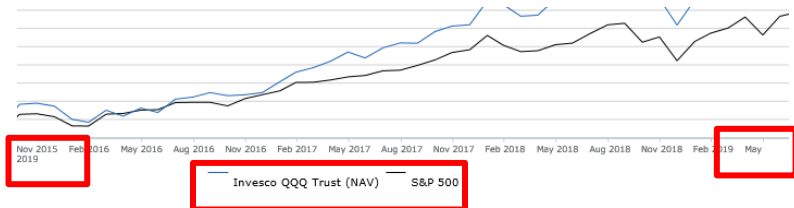


# Chart Labels in Wrong Position

PDF

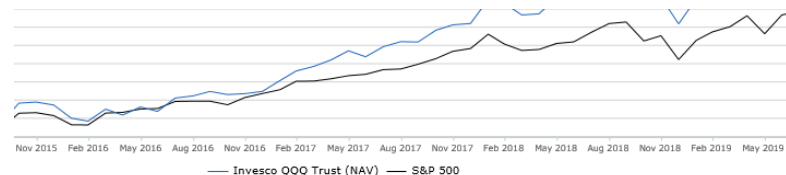


12246



Some labels are incorrect

12602



Conversion is correct

Original name

SHA1

46ae0f58b12e857228044f65f390325214b3005f

FB

23128

TESTED

# Reconstruction as Excel



# Spurious Blank Line Created at Page Break

PDF

18	43		10,23,98,588
19	44		10,23,98,588
20	45		10,23,98,588

21	46		10,23,98,588
22	47		10,23,98,588
23	48		10,23,98,588

Table is split over page boundary

12246

19	18	43		10,23,98,588
20	19	44		10,23,98,588
21	20	45		10,23,98,588
22				
23	21	46		10,23,98,588
24	22	47		10,23,98,588
25	23	48		10,23,98,588

A blank row is present

12602

19	18	43		10,23,98,588
20	19	44		10,23,98,588
21	20	45		10,23,98,588
22	21	46		10,23,98,588
23	22	47		10,23,98,588
24	23	48		10,23,98,588

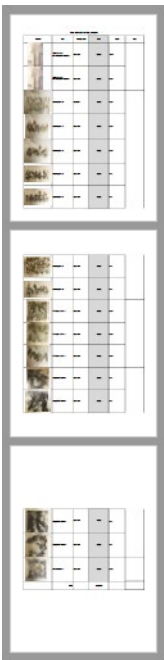
Blank row is no longer created

Original name	
SHA1	835b86bc02278b9a6a6f78f3962c45bdb5dce731
FB	24876



# Table Detection Improved

PDF



12246

PDF has a table split over three pages

12602

Data has only been extracted from first two pages

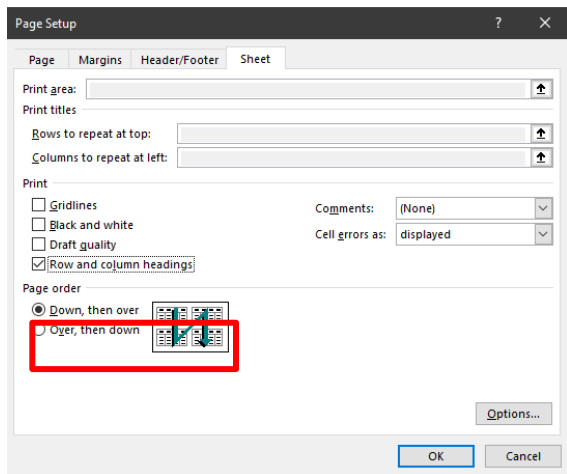
All of the data has been extracted as a single logical table

Original name	
SHA1	1b11872c7a4c3971e3c053e9924e8f578c8581bc
FB	27682



# Printing Artifacts

Excel allows Row and Column Headings to be included when creating a PDF



This results in a PDF that includes information that was not part of the original table.

In some cases, Solid Framework now automatically removes these artifacts.

In a survey of more than 20,000 PDFs collected from a range of sources, this affects about 1 in 1000 files

# Printing Artifacts Correctly Removed - Example 1

PDF

	A	B	C
1	City	Population	Area (km2)
2	Auckland	1,470,100	607.10
3	Christchurch	383,200	295.15
4	Wellington	215,100	112.29
5	Hamilton	176,500	110.37
6	Tauranga	151,300	135.12
7	Lower Hutt	110,700	78.52
8	Dunedin	106,200	91.58
9	Palmerston North	81,500	76.92
10	Napier	66,300	104.90
11	Hibiscus Coast	59,800	42.90
12	Porirua	59,600	60.96
13	Rotorua	58,500	48.12
14	New Plymouth	57,600	75.49
15	Whangārei	54,400	57.06
16	Nelson	51,100	54.33
17	Invercargill	50,200	60.70

Row and column headers have been exported, but are not really part of the data

12246

A	B	C	D	
1	A	B	C	
2	City	Population	Area (km2)	
3	2	Auckland	1,470,100	607.10
4	3	Christchurch	383,200	295.15
5	4	Wellington	215,100	112.29
6	5	Hamilton	176,500	110.37
7	6	Tauranga	151,300	135.12
8	7	Lower Hutt	110,700	78.52
9	8	Dunedin	106,200	91.58
10	9	Palmerston North	81,500	76.92
11	10	Napier	66,300	104.90
12	11	Hibiscus Coast	59,800	42.90
13	12	Porirua	59,600	60.96
14	13	Rotorua	58,500	48.12
15	14	New Plymouth	57,600	75.49
16	15	Whangārei	54,400	57.06
17	16	Nelson	51,100	54.33
18	17	Invercargill	50,200	60.70

The row and column headers have been extracted

12602

	A	B	C
1	City	Population	Area (km2)
2	Auckland	1,470,100	607.10
3	Christchurch	383,200	295.15
4	Wellington	215,100	112.29
5	Hamilton	176,500	110.37
6	Tauranga	151,300	135.12
7	Lower Hutt	110,700	78.52
8	Dunedin	106,200	91.58
9	Palmerston North	81,500	76.92
10	Napier	66,300	104.90
11	Hibiscus Coast	59,800	42.90
12	Porirua	59,600	60.96
13	Rotorua	58,500	48.12
14	New Plymouth	57,600	75.49
15	Whangārei	54,400	57.06
16	Nelson	51,100	54.33
17	Invercargill	50,200	60.70

Only the actual data has been extracted

Original name

SHA1

04e0e3cc33a978ecf822bcbf487f8822c4b2e940

FB

27369

TESTED



# Printing Artifacts Correctly Removed - Example 2

PDF

	A	B	C	D
1				
2				
3				
4			PERIODO	1 de febr
5				
6				
7	CEDULA	NOMBRE	CARGO	SUELDO
8				
9			ANALISTA SISTEMAS	2.500,00
10			AUX.CONTABLE	700,00
11			CONTADOR	1.800,00
12			DIGITADOR	545,00
13			GERENTE	3.500,00
14			MENSAJERO	566,67
15			OPERARIO BODEGA	600,00
16			SECRETARIA	800,00

Row and column headers have been exported, but are not really part of the data

12246

	A	B	C	D	E
1		A	B	C	D
2	1				
3	2				
4	3				
5	4			PERIODO	1 de febr
6	5				
7	6				
8	7	CEDULA	NOMBRE	CARGO	SUELDO
9	8				
10	9			ANALISTA SISTEMAS	2.500,00
11	10			AUX.CONTABLE	700,00
12	11			CONTADOR	1.800,00
13	12			DIGITADOR	545,00
14	13			GERENTE	3.500,00
15	14			MENSAJERO	566,67
16	15			OPERARIO BODEGA	600,00
17	16			SECRETARIA	800,00

The row and column headers have been extracted

12602

	A	B	C	D	E
1					
2					
3					
4			PERIODO		1 de febr
5					
6					
7	CEDULA	NOMBRE	CARGO	SUELDO	
8					
9			ANALISTA SISTEMAS	2.500,00	
10			AUX.CONTABLE	700,00	
11			CONTADOR	1.800,00	
12			DIGITADOR	545,00	
13			GERENTE	3.500,00	
14			MENSAJERO	566,67	
15			OPERARIO BODEGA	600,00	
16			SECRETARIA	800,00	

Only the actual data has been extracted

Original name	
SHA1	b395e6a1854bcff2c34c6d787e8714954961868b
FB	27369

TESTED

# NSE Improvements

We have continued to improve conversion of PDFs that contain Non Standard Encoding.

The following slides are just a few examples of the huge number of improvements.

# Improved NSE in Cyrillic Document

PDF

Движения клеток также  
миозин, динеин). Белки  
из клетки и внутри к  
ляторными выполняют  
ками. И наконец, белки  
Углеводы подраздел

The document is encoded  
in a non-standard way

12246

Движения клеток также  
миозин, динеин). Белки  
из клетки и внутри к  
ляторными выполняют  
ками. И наконец, белки  
Углеводы подразде

Several characters are  
incorrect

12602

Движения клеток также  
миозин, динеин). Белки  
из клетки и внутри к  
ляторными выполняют  
ками. И наконец, белки  
Углеводы подраздел

Conversion is correct

Original name	
SHA1	1446897d44000463d34f0c3553c1981db37c0045
FB	27762

TESTED

# NSE of lookalike characters is now correct

PDF

Il flusso di calore che passa dall'interno dell'edificio verso l'esterno (da T

The text is encoded in a non-standard way. In this font, “I” and “l” look the same but should be treated differently

12246

Il flIusso di caIore che passa daII interno deII'edificio verso I'esterno (da T

Export as text shows that some of the characters are incorrect.  
(Issue exists in Word but is difficult to see)

12602

Il flusso di calore che passa dall'interno dell'edificio verso l'esterno (da T

Conversion is correct

Original name	
SHA1	c60ddfb52776dda18f157e89146bf463aacff697
FB	27749



# Other Improvements

FB26411: Improved table detection when including non-table content

# Number of FogBugz Cases with Automated Tests

	30th Nov 2020	12 <sup>th</sup> Aug 2021	6 <sup>th</sup> Oct 2021
Regression cases - Active	c. 210	311	316
Regression cases - Closed	1155	1991	2026
All cases - Active	c. 1000	1301	1370
All cases - Closed	2590	4118	4265
Total number of tests	c. 3600	5419	5635

A huge effort has gone into creating automated tests to ensure that things that are fixed *stay* fixed.

In addition there are more than 80,000 other tests which are run on multiple machines and OSs on every build.