Understanding Document Rotation

Getting the Most from Solid Framework

3rd July 2019

Solid Framework is an awesome tool for reconstructing documents from PDFs.

It can deal with both scanned and "Saved as" PDFs

Solid Framework tries to help you

It is common when dealing with scanned documents to find that:

- some pages were scanned upside down, and others the right way up
- some pages were portrait and others were landscape format



This can also happen with "Saved As" files that have been modified in a PDF editor

Solid Framework automatically detects the *dominant* text direction, and rotates the page so that text is horizontal



alender	190
ne lend	Jan 27 May Hall to Ale, Child Beiry dan.
and Pel. WorkProcessor Meeting, International researces, Hollewsk, MA, Style of Re-markh in All	Pub? Pd West Processor Meeting, Lonation 1983 Style of the month in Reser Reed Contact July Flowed, pdfDaward.org
nigan Alani Caniani Jell Pland, pelitjevel og nr 7. Bal: Westen Dit Hansfore Competition, mediali, MA: A SIGBOOT word Caniadi Cham « Calassenis, 43 773-5352 m 1. Wed: Fourth Asstal Direy Yan's Day Jonal	Public Field Could Britch Water Beer Period Olingow, Social Contact CAMA
	Pub 22 Ani 3rd Assend Rosine Hanslews Compu- bilion, Isoation THE: A NUMBERTY scenii. Cardani Frin Langhin, print@work.org.
ini, Spinnele, Basine - BYOH & HYOHYA	Nor 10 Non Lucions Lagor, Chile Roley due.
er3 M Wet Process Meeting London 8D. Syle of the month is Pub Alm & SPAd Com-	Res 77 West Associationary Party, Databa 1966. Reg 77 Ded Associal West Marylini, Databa
en del France, pergenerarg en 18 fei - Nieth Annual Statione New Regioni aginal Humderer Competition Collin Mill Sector Index, CT. A 10000/CTT ment Costant Orig	THA Jone 31A22 / Jone 3628 Third Annual World Hencinew Context, Failurer College, Context Devic Soluble, Reales New Concessor
law, Oragificia company and	
n 77 Bel Second Annual West Processor Ratio in & Dele Deves Learning Linear Station, Delettin	
handines yaitad classificati	
has been an the continge, which man provide the	
na, bad i 'E unad your belp.	Dues Are Due
he Wate have always given a helping hand when deal, and now Proceeding. Public to build exceeding	Duco nic Duc
18 PM, December 11, a Westweder, at the West-	
in to damage some of the things that we will do for in and your, and got a secon of who can bely out.	\$ £ ¥
Use to back out meeting backing, severables with m, activities, and all of the flat staff. that makes the	ΨΔΤ
his talk. If you're new to the ability a have wanted to income with 8, this is a great shares to get involved.	Denision and
lease some, participate, and help make the shik- reschedue!	Pay up now!
he had note 1 could() fit in the Transver's report, at we have \$21.00.77 and only %40 reasonal to date.	
	Volume 12, Number 13

here is ships have readen of the later.	ing potions for a bial of \$25 (including a grand lig)	and position filled with training there asses the horizony time. The Automatic terrory, located on a manip mail in the colling	a jabe its me noti vio dat adaily ince levering ded)
If you could induce, five you may have lower labout New Street, a Stelator law bay in	I date 's easier Horasotte too late a night along the sent day I wanted to be an early to get any wedge are	Archester, & a set beware Con- viewind this deads, they have a reacting start has that dealing an	Unite Louis' devotors, he doo- blad Chinesy's served output to 100,000 heardines. He assess
Reauty will many as tap. Well, not appears. I want to 119 Mind Asymptotic and Imagination design relationst will be price. You are will find-control on to deal	and drive to the Archemen Lorent 1 half an invitation to the Research D'Archemile and 1 drive's reset to miss 6. Xented and in Theory are much to research in Theory and	a lotte and standard and fe- mention (glycel political). The child part of the tenency in their pain call, originally from the may faint over has a secondard	pialed the using production on tool optimals that allowed him to absorb or production without
team Mini Sulvin nan an Alexador 20's style har and the Orand Plane har many ora	Here between and per set a deal for \$112 one time days with all because administration argues (yes	and discission is in Adv rates	using additional latter. This is what being life him to the Automati- ber way. Constant works to dow the force sectors when the other
For E25, Donghi a few sugar	HE2) One is pressing, but basis in that part of the scrattey do not serve usually as much as pro-	Also from Chiney is one of Seir Inverse, Louis Front, Louis Ine worked in the Polgian Inversion	neuris le voiles
with every data	made if you want to get in the for-	of these he agend at Chiney He	To this weig Cheinian is prepared to speed the money that it islam.
Shill, my vale weat to in '1 Spin- relayite at 1 Plane Do Jacks (Darden Spane for an New	-	exprine Association Loss, the mode of Danay do very	electronic fait significantly access in efficiency. They are
Reglamber). Here Day have great been on lay and in boths, plan here star main universe. I had a three scores med, have with	I have to say up front that Obta- tion Decembers, owner (howal) title Develop), and that he been particle one well as mandating	many or one ways. They have the table, but they generally have the work to have bedy. Loais was one such himd have, and he table on that he was the hard become in	mercedly treating a just in the pand production factory flat, of the pash of a batter, grown pand others and a compute agent into the pand of the bars films of
stay and generative, and a	and I'm a happy maketing loof.	presiden, Europh and in hills,	year early by here time (they

Rotation is on a page by page basis



Sample file can be found at https://solidframework.net/wp-content/uploads/samples/textbox_orientation_multipage.pdf

This conversion is obvious when converting to Word, but the same effect occurs when a "Core Model" is created.

The Core Model is a structured "in-memory" representation of the PDF contents and offers many opportunities to deal with the *data* within the PDF without the need to convert to a Word Document.

Is this behaviour optional?

You can prevent the Converter from rotating a page left or right by 90 degrees (changing landscape to portrait or vice versa) using

converter.AutoRotate = false

It is *not* possible to prevent pages being rotated by 180 degrees if the dominant text is upside down.

It is *not* possible to prevent this behaviour when using the Core Model.

Identifying that page rotation occurred

GetPageWasAutoRotated(pageNumber) indicates the angle (in degrees clockwise) that the page was rotated in order to get it "the right way up"



Note: "pageNumber" starts at one for the first page – even if a subset of pages were used to create the Core Model

What happens if there are different text orientations on the same page?

Solid Framework will detect the dominant text direction and rotate the page so that text is the right way up.

This can be a little confusing!

Dealing with different text orientations on the same page

The sample file contains fours pages

- each page has same general layout,
- one text box on each page contains more text which will make it dominate



Sample file can be found at https://solidframework.net/wp-content/uploads/samples/rotated_textboxes_landscape.pdf

Dominant text is right way up



Word

No change

Dominant text is upside down



Page rotated upside down

Dominant text runs from top to bottom



Page rotated left

Dominant text runs from bottom to top



Page rotated right

Identifying the orientation of text in the original PDF

Only applies to text within Text boxes

Logic for calculating the direction of text in the original document is complex.

Need to use:

Property	Provides information about
textbox.Rotation	If a text box has been rotated by 180 degrees
textbox.TextDirection	If the text is horizontal or vertical
GetPageWasAutoRotated(pageNumber)	If the page was rotated so that the dominant text is horizontal

Sample Code

```
private TextBoxRotation GetRotationForTextBox(SolidFramework.Model.Plumbing.TextBox tb, LayoutDocument ld, CoreModel cm)
```

```
int workingRotation = 0;
```

{

}

```
//the TextBox rotation is only really useful to identify whether the textbox has been rotated 180 degrees, since vertical text does not require the text box to be rotated.
if (tb.Rotation > 0)
```

```
//Ensure the rotation is right-angled (divisible by 90), since other values are possible. This is most likely if the document was scanned.
workingRotation = (int)(Math.Round((tb.Rotation / (double)90)) * 90) % 360;
```

```
//TextDirection.Horizontal may relate to text that was originally the right way up, or which was upside down.
```

```
if (tb.TextDirection != TextDirection.Horizontal)
```

```
switch (tb.TextDirection)
        case TextDirection.Rotate90:
            workingRotation = 90;
            break;
        case TextDirection.Rotate270:
            workingRotation = 270;
            break;
        default:
            break;
    }
}
int pageNumber = -1;
int pageDir = -1;
int initialRotation = -1;
int finalRotation = -1;
//Get the LayoutTextBox that contains layout information for the textbox. This is done as one method of getting the pagenumber
LayoutTextBox ltb = ld.FindLayoutObject(tb.GetID()) as LayoutTextBox;
if (ltb != null)
    pageNumber = ltb.GetPageNumber();
    pageDir = cm.GetPageWasAutoRotated(pageNumber);
    initialRotation = (360 + workingRotation - pageDir) % 360;
    finalRotation = workingRotation;
}
```

return new TextBoxRotation(tb.TextDirection, tb.Rotation, initialRotation, finalRotation, pageDur, pageNumber);

}

Summary

Solid Framework aims to minimise the number of choices that users have to make by setting logical defaults.

For most users this works well.

If required, the orientation of text can be calculated.