Cool.

How do you handle when the text you're extracting from the file doesn't match the order it appears on the rendered page?