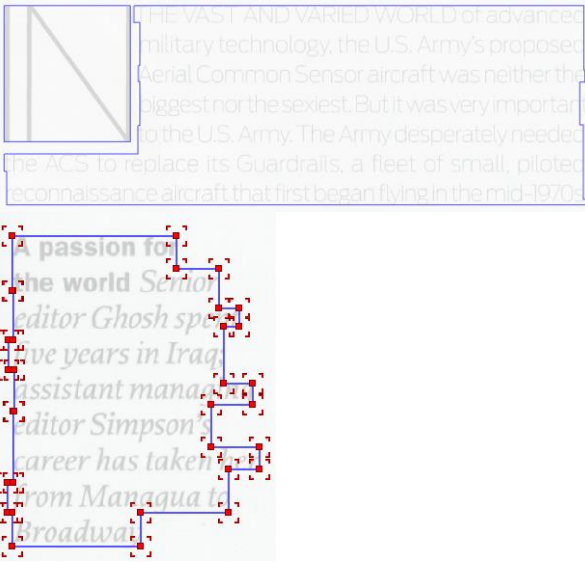


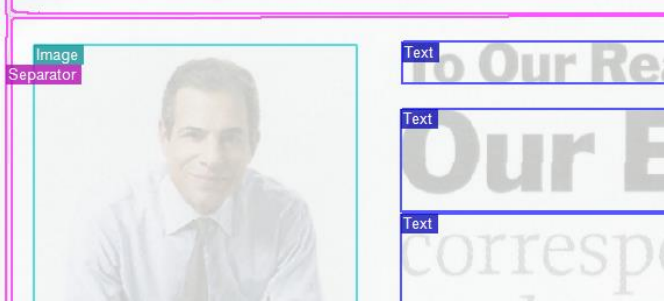

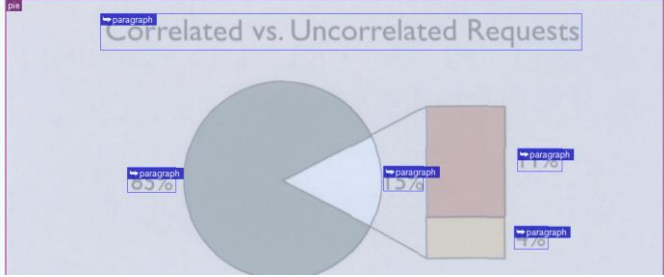
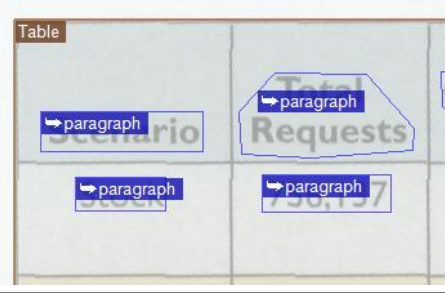
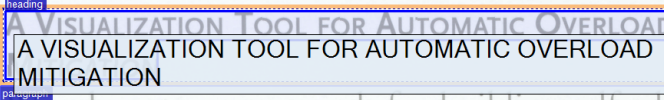
# ICDAR 2017 Competition on Recognition of Documents with Complex Layout – RDCL 2017

## Examples

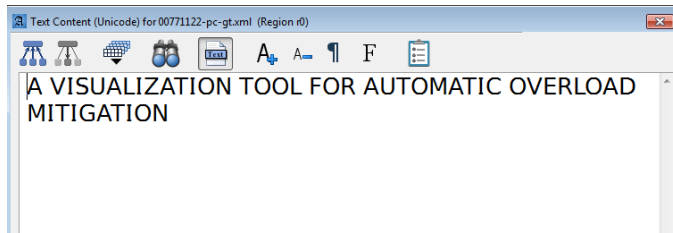
In this document we show a few examples of what we ask participants of the competition to submit. We also indicate the importance of the different aspects:

- High priority: This is essential content – concentrate on this data
- Low priority: Less important for winning the competition, but still evaluated
- Optional: Not evaluated, but might be mentioned especially in the competition paper

Type of Data	Example	Notes	Priority															
Region outlines		<ul style="list-style-type: none"> <li>• Polygons</li> <li>• Ideally closely fitting</li> <li>• No overlap</li> <li>• For text: Capture a whole paragraph; never across columns; drop capitals as separate regions</li> </ul> <p>XML example:            &lt;TextRegion ...&gt;            ...            &lt;Coords            points="19,2373            19,2959 6,2959            6,2373"/&gt;</p>	High priority															
Main Region Types	<p>Type</p> <table border="1"> <tr> <td><input checked="" type="checkbox"/> Text</td> <td><input type="checkbox"/> Image</td> <td><input type="checkbox"/> Graphic</td> </tr> <tr> <td><input type="checkbox"/> Line Drawing</td> <td><input type="checkbox"/> Table</td> <td><input type="checkbox"/> Chart</td> </tr> <tr> <td><input type="checkbox"/> Separator</td> <td><input type="checkbox"/> Maths</td> <td><input type="checkbox"/> Chem</td> </tr> <tr> <td><input type="checkbox"/> Music</td> <td><input type="checkbox"/> Advert</td> <td><input type="checkbox"/> Noise</td> </tr> <tr> <td><input type="checkbox"/> Unknown</td> <td></td> <td></td> </tr> </table>	<input checked="" type="checkbox"/> Text	<input type="checkbox"/> Image	<input type="checkbox"/> Graphic	<input type="checkbox"/> Line Drawing	<input type="checkbox"/> Table	<input type="checkbox"/> Chart	<input type="checkbox"/> Separator	<input type="checkbox"/> Maths	<input type="checkbox"/> Chem	<input type="checkbox"/> Music	<input type="checkbox"/> Advert	<input type="checkbox"/> Noise	<input type="checkbox"/> Unknown			<ul style="list-style-type: none"> <li>• Most important is the differentiation between text and non-text</li> <li>• <b>Yes:</b> Text, image,</li> </ul>	High Priority
<input checked="" type="checkbox"/> Text	<input type="checkbox"/> Image	<input type="checkbox"/> Graphic																
<input type="checkbox"/> Line Drawing	<input type="checkbox"/> Table	<input type="checkbox"/> Chart																
<input type="checkbox"/> Separator	<input type="checkbox"/> Maths	<input type="checkbox"/> Chem																
<input type="checkbox"/> Music	<input type="checkbox"/> Advert	<input type="checkbox"/> Noise																
<input type="checkbox"/> Unknown																		

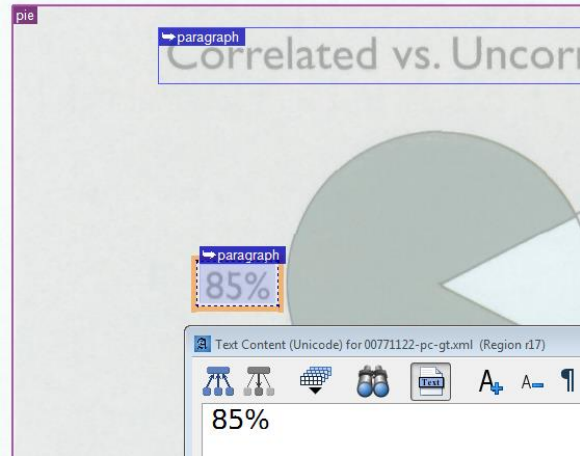
Type of Data	Example	Notes	Priority
		<p>graphic, table, chart, maths, separator, noise</p> <ul style="list-style-type: none"> <li>• <b>No:</b> no advert, no chemical notation, no musical notation</li> </ul> <p>XML example:            &lt;TextRegion ...&gt;            &lt;ImageRegion ...&gt;</p>	
<b>Subtypes</b>		<ul style="list-style-type: none"> <li>• Part of the ground truth, but not required</li> </ul> <p>XML example:            &lt;TextRegion type="paragraph" ...&gt;</p>	Optional
<b>Nested Text Regions</b>	 <p>Pie chart with nested text regions</p> 	<ul style="list-style-type: none"> <li>• Only for readable and front-facing text</li> <li>• E.g. labels in graphics, scene text, table cells</li> </ul> <p>XML example:            &lt;ChartRegion ...&gt;            &lt;TextRegion ...&gt;            &lt;/TextRegion ...&gt;            &lt;/ChartRegion ...&gt;</p>	High priority
<b>Main text content</b>		<ul style="list-style-type: none"> <li>• OCR text for each (non-nested) text region</li> </ul>	Low priority

Type of Data	Example	Notes	Priority
--------------	---------	-------	----------



- As seen on image
- XML example:  
<TextRegion ...>  
  <TextEquiv ...>  
    <Unicode>  
...

**Text content of nested regions**



Not evaluated, definitely not penalised if missing.

Optional