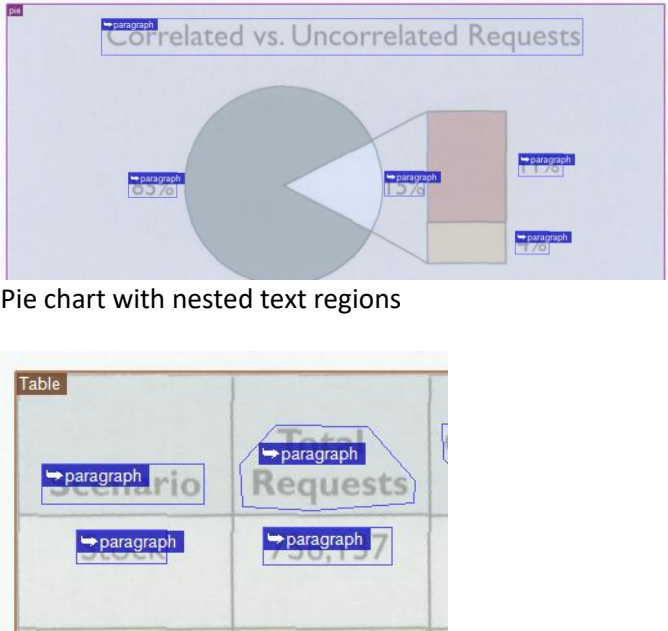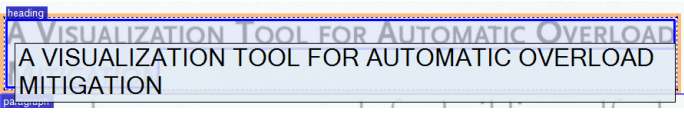# ICDAR 2019 Competition on Recognition of Documents with Complex Layout – RDCL 2019
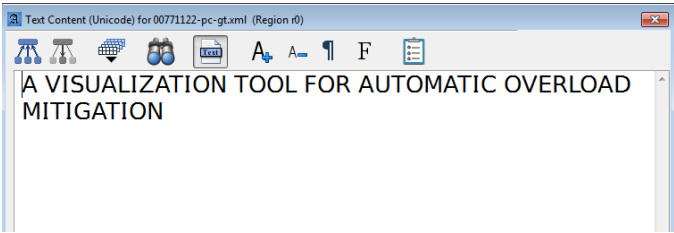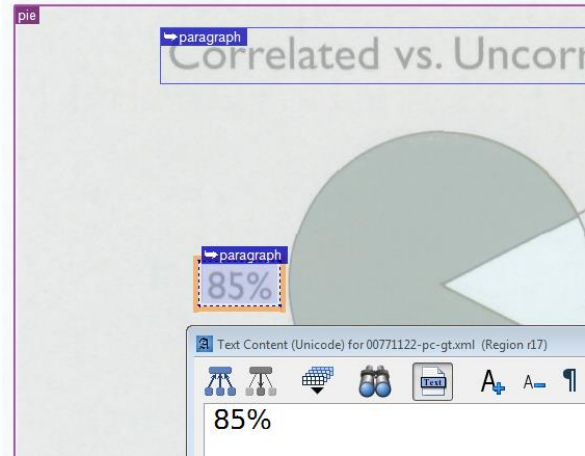
Examples

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><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>XML example:<br>`<TextRegion …>`<br>`…`<br>`<Coords points="19,2373 19,2959 6,2959 6,2373"/>` | High priority |
| **Main Region Types** |  | <ul><li>Most important is the differentiation between text and non-text</li><li>**Yes:** Text, image,</li></ul> | High Priority |

| Type of Data | Example | Notes | Priority |
|---|---|---|---|
| |  | graphic, table, chart, map, maths, separator, noise<br>• **No:** no advert, no chemical notation, no musical notation<br><br>XML example:<br>`<TextRegion …>`<br>`<ImageRegion …>` | |
| **Subtypes** |  | • Part of the ground truth, but not required<br><br>XML example:<br>`<TextRegion`<br>`type="paragraph"…>` | Optional |
| **Nested Text Regions** | <br>Pie chart with nested text regions<br><br> | • Only for readable and front-facing text<br>• E.g. labels in graphics, scene text, table cells<br><br>XML example:<br>`<ChartRegion …>`<br>`  <TextRegion …>`<br>`  </TextRegion>`<br>`</ChartRegion>` | High priority |
| **Main text content** | <br>A VISUALIZATION TOOL FOR AUTOMATIC OVERLOAD MITIGATION | • OCR text for each (non-nested) text region | Low priority |

| Type of Data | Example | Notes | Priority |
|---|---|---|---|
| |  | • As seen on image<br><br>XML example:<br>`<TextRegion …>`<br>`  <TextEquiv …>`<br>`    <Unicode>`<br>… | |
| **Text content of nested regions** |  | Not evaluated, definitely not penalised if missing. | Optional |