Table Regions in PAGE XML

<TableRegion>

# General Attributes (since PAGE 2009)

* Orientation
* Rows (int)
* Columns (int)
* Line separators (True/False)
* Line colour
* Background colour
* Embedded text

# Nested Regions (since PAGE 2016)

A table region can have child regions of any type, representing table cells.

# TableCellRole (since PAGE 2017)

Extension for a nested region that is a table cell.

Attributes:

* rowIndex
* rowSpan
* columnIndex
* colSpan

# Table Grid (PAGE 2018?)

* Physical location of the table grid lines (visible or invisible)
* Can be used to define the cell spaces (e.g. for further processing)
* Grid of points, size = (rows+1) \* (columns+1)

# Comments

* Consistency between rowIndex+span / columnIndex+span and region coordinates should be ensured by the application (e.g. Aletheia)
* Non-rectangular grid structures could be done via overlapping rowIndex+span / columnIndex+span areas, but should be discouraged for now
* In future we can think about cell border attributes

Example:

Grid (with points at junctions)

Nested region

rowIndex = 0, rowSpan = 1

columnIndex = 0, colSpan = 2

TableRegion (rows = 6, columns = 5)

Nested region

rowIndex = 2, rowSpan = 2

columnIndex = 2, colSpan = 3