Text2Image Parameters

thresh

```
          Threshold for text alignment. If the confidence of a text-to-ima
ge
           alignment above this threshold, an alignment is done (default =
0.0). A
            good value is between 0.01 and 0.05. Note that the confidence
is stored
            in the pageXML anyway, so deleting text alignments with low co
nfidence
            can also be made later.
```

hyphen

can be null, a non-negative double value or a json-string (default:
null). If no value is set or value is “Infinity”, no hyphenation is done.
If value is a positive double value, the value are the additional costs
to recognize a hyphenation. The default hyphenation signs at the end of
the line are ‘¬’, ‘-‘, ‘:’, ‘=’. The default hyphenation signs at the
beginning of the line are empty. There can be hyphenations between all
letter-pairs. If one wants to use hyphenation rules for a specific
language, this can be configured using the key ‘hyphen_lang’.
The hyphenation sign in the groundtruth will be ‘¬’.
If one wants to configure more, one has to write a j-son-string.
Keys:
prefixes: list of hyphenation sign that can be hyphens at the
beginning of a line (default: empty)

suffixes: list of hyphenation sign that can be hyphens at the end of
a line (default: empty)

skipSuffix: boolean if suffix is optional (true) of forced (false)
(default: false)

skipPrefix: boolean if prefix is optional (true) of forced (false)
(default: false)

hypCosts: non-negative value that produces additional costs to recognize a hyphenation. (default: 0.0)

pattern: language pattern (e.g. EN_GB, EN_US, DE, ES, FR,…) (default: empty)

example: “{

“skipSuffix”:false,

“skipPrefix”:true,

“suffixes”:[“¬”,”-“,”:”,”\u003d”],

“prefixes”:[“:”,”\u003d”],

“hypCosts”:6.0,

“pattern”:”EN_GB”

}”

one of the 4 suffixes have to be recognized and one of the both prefixes can be recognized. Hyphenation costs of 6.0 are added. Hyphenation is only possible as defined for language EN_GB.

hyphen_lang

```
    if hyphen is given, hyphenation-rules from different languages can be
    applied. If value = null or empty, a linebreak between all letters is
    possible (unicode-characters of Category L). Otherwise, a rule is applied
    ( see https://github.com/mfietz/JHyphenator.git for details). The
    language have e.g. "DE" for German and "EN" for English. Default = null.
```

skip_word

```
    makes it possible to skip a word, for example if a baseline is too sho
rt
    (default: null). The value have to be a positive double value. It
    repesents the default delete-costs for each character. A good value is
    4.0. The higher the value, the less words were skipped. If value = 0,
a
    word can be deleted without producing costs (destroys the algorithm),
if
    value = Infinity, no characters can be deleted.
```

skip_bl

```
    makes it possible to skip a baseline (default: null). Sometimes the LA
    finds a baseline in noise (aka false positive). It is possible to dele
te
    those baselines instead of "pressing" a sequence into the line. The va
lue
    has to be positive double value. The lower the value, the easier a lin
e
    is ignored. A good value is 0.2.
```

jump_bl

```
    makes it possible to handle wrong reading order in the LA (default: nu
ll)
    The value makes it possible to jump instead of the after a line to eve
ry
    other line. If value = 0, the reading order has no effect at all. If
    value = Infinity is the same like value = null. If you cannot trust th
e
    reading order, set value = 0.
```

best_pathes

```
    if the number of confmats and references gets too large, one can only
    keep a specific number of paths at each reference. As default all path
s
    are calculated (like setting value = Infinity). A good value is 200.0
```