

do qual to do a apresentar um resultado final semelhante e com capacidade de armazenamento de meta-
dados superior por fim, sendo baseado em XML, torna sua manipulação mais complexa e, em múltiplas
casos a informação é atribuída a classes de blocos e subtítulos). Assim sendo, embora tenha sido desenvolvido um conversor de, e para HOCR,
se pela criação de uma estrutura de dados própria.

do, foi implementada uma estrutura de dados no formato de árvore de dados.

==>

Level
(page|block|par|line|word)_num

text

conf

id

type

children

box

Descrição:

Argumentos:

Descrição:

Argumentos:

Descrição:

Argumentos:

Descrição:

Argumentos:

Descrição:

Algoritmo:

Argumentos:

Descrição:

Algoritmo:

$box.width \geq$

$box.height * 4 || box.height \geq$

$box.width * 4$

Argumentos:

Descrição:

Algoritmo:

Argumentos:

Descrição:

Argumentos:

Descrição:

Argumentos:

Algoritmo:

//

Linha

única

//

Palavra

única

//

Múlti-

plas

palavras

//

Ver-

ifica

se

a

maio-

ria

das

palavras

co-

m-

ci-

dem

hor-

zon-

tal-

mente

//

Múlti-

plas

lin-

has

//

Ver-

ifica

se

a

maio-

ria

das

lin-

has

co-

m-

ci-

dem

hor-

zon-

tal-

mente

Descrição:

Argumentos: