

Original Data

1. Example

Here is an example.

As shown in Table 1.

1	2
3	4

PNG

Source-code

LaTEX Source-Code

Stage-one: Data Preprocessing

`\input{method}`
`\input{related}`
`\input{intro}`

Expand all referenced files

~~`\vspace`~~ ~~`\ref`~~
~~`\textbf`~~ ~~`comments`~~

Remove redundant tags that have no semantic information

`.pdf`
`.eps` `.jpg` → `.png`

Standardize figures to PNG format

`hyperref` → `hyperref`

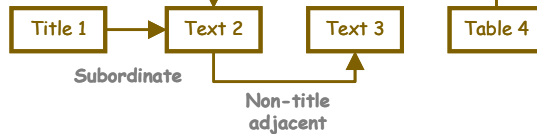
Remove color attribute of 'hyperref'

Stage-two: Units Segmentation

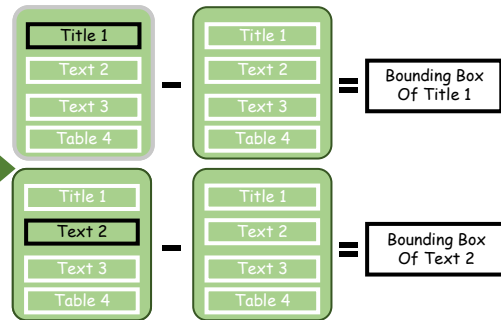
```
[  
  {'section': '\\section{Example}'}, 1  
  '\\n'.  
  'Here is an example.', 2  
  '\\n'.  
  'As shown in Table \\ref{table}.', 3  
  '\\n'.  
  {'tabular': '\\begin{tabular}{|c|c|} \\n 1 & 2 \\ \\ \\hline \\n 3 & 4 \\ \\ \\hline \\n end{tabular} \\n \\label{table}'}, 4  
  '\\n'  
]
```

Stage-three: Attribute Assignment and Relation Retrieval

Explicitly-referred



Stage-four: Color Rendering



★ Meta Data

