Combinación de texto e información visual

Texto plano

Texto plano extraído con pypdf

Información visual

Texto extraído con OCR y reconstruido

Separar texto en palabras (txt_words)

Separar texto en palabras (ocr_words)

Obtener lista de diferencias con difflib

No diferencia

```
Capítulo
Capítulo
```

```
Capítulo
```

Palabras cortadas

```
- Constituciona
+ Constituciona-
?              +
- -
```

```
Constituciona-
```

Palabra diferente

```
- II.      - 5.
?  ^
+ IT.     + s.
?  ^
```

```
II.        5.
```

Múltiples palabras diferentes

```
- Orgánica
- de
- la
+ ORGÁNICA
+ DE
+ LA
```

```
Orgánica
de
la
```

Palabra separada

```
- DÍA
+ D
+ Í
+ A
```

```
D
Í
A
```

Una palabra sin correspondencia

```
- 1.
La
```

```
1. La
```