

# BOEs-NLP

Commits:

- Separar boletines en scripts Python separados
- Separar búsqueda de Tags en BBDD ASECORP en script separado
  - Hacer que se compruebe modificación de la BBDD, y si no se modifican Tags encontrados no se ejecuta de nuevo el Script
  - Buscar patrones REGEX en filas en las que no se han encontrado Tags
- Incluir extracción Tags en contenido PDFs
- Evaluar librería **Schedule**

## BOE

- Sólo considera Sección I. Disposiciones generales
- Intentar incluir también Sección Anuncios, con código similar a :  
`el = doc.xpath("//div[contains(@class, 'channel') and not(contains(@class, 'disabled'))]")`

## DOGC

- No implementado [\[1\]](#)

## DOUE

- De momento sólo incorpora sección L

## DOCM (Diario Oficial de Castilla la Mancha)

- Incorpora todas las secciones

- Considerar eliminar algunas secciones

1. This is a footnote. ↩