



PRÁCTICA 1

Tipología y Ciclo de Vida de los Datos



10 DE NOVIEMBRE DE 2025

ANGEL BARRÓN ALMENDROS Y PEDRO JOSÉ BERBERANA MARTÍN

UOC

[Escriba aquí]

ENLACES

Recurso	Descripción	URL
Web	Página oficial del Consejo General del Poder Judicial.	https://www.poderjudicial.es
Repositorio	Código fuente y scripts de extracción del dataset.	https://github.com/UOC-Tipologia-y-ciclo-vida-datos/UOC-PRA1-Juzgados-de-Espanna
Zenodo	Publicación y DOI del dataset generado.	https://zenodo.org/records/17570935 DOI: 10.5281/zenodo.17570935
Vídeo	Presentación audiovisual del proyecto de práctica	https://drive.google.com/file/d/1T12bT5BxXmefrQWfS2-B7VJDk-JMMVeE/view?usp=sharing

1 CONTEXTO

El proyecto se realiza con el objetivo de recolectar y estructurar la información del sitio web oficial del Consejo General del Poder Judicial (CGPJ), un sitio público accesible a través de esta web <https://www.poderjudicial.es>.

Se trata de una fuente fiable, ya que es un portal institucional del gobierno español. Los datos publicados se actualizan y son verificados por el propio CGPJ con el objetivo de mantener la transparencia y permitir el análisis.

El proyecto se centra en la recopilación de información sobre los juzgados españoles, incluyendo su tipología, ubicación, datos de contacto, nombre del juez y otros datos que resulten relevantes.

2 TÍTULO

Jueces y Juzgados de España – Datos públicos del Consejo General del Poder Judicial.

3 DESCRIPCIÓN DEL DATASET

El dataset obtenido recopila información pública sobre los juzgados españoles a partir de la web del CGPJ. Esta información se extrae a partir de técnicas de web scraping en Python.

Cada registro obtenido representa cada juzgado de manera individual, pudiendo discernir sus diferentes datos y características (localización, tipología, datos de contacto...). De esta manera, este dataset permite el análisis de los distintos órganos judiciales.

4 REPRESENTACIÓN GRÁFICA

A continuación, se representa el recorrido que realiza el código de web scraping a través de la web y los datos que recopila:

- Desde el [Directorio de Órganos Judiciales](#), se captura cada una de las provincias y la URL que enlaza al listado de los juzgados de esa provincia:

Estás en: Directorio de Órganos Judiciales

Directorio de Órganos Judiciales

Selecciona tu provincia y municipio y encuentra el Juzgado que estás buscando.

The screenshot shows a map of Spain with provincial boundaries. To the right of the map is a vertical list of Spanish provinces, each preceded by a right-pointing arrow. The list is enclosed in a red rectangular border. The provinces listed are: A Coruña, Albacete, Alicante, Almería, Araba/Álava, Asturias, Ávila, Badajoz, Barcelona, Bizkaia, Burgos, Cáceres, Cádiz, Cantabria, Castellón/Castelló, Ceuta, Ciudad Real, Córdoba, Cuenca, Gipuzkoa, Girona, Granada, Guadalajara, Huelva, Huesca, Illes Balears, Jaén, La Rioja, Las Palmas, León, Lleida, Lugo, Madrid, Málaga, Melilla, Murcia, Navarra, Ourense, Palencia, Pontevedra, Salamanca, Santa Cruz de Tenerife, Segovia, Sevilla, Soria, Tarragona, Teruel, Toledo, Valencia/València, Valladolid, Zamora, and Zaragoza.

Provincia
A Coruña
Albacete
Alicante
Almería
Araba/Álava
Asturias
Ávila
Badajoz
Barcelona
Bizkaia
Burgos
Cáceres
Cádiz
Cantabria
Castellón/Castelló
Ceuta
Ciudad Real
Córdoba
Cuenca
Gipuzkoa
Girona
Granada
Guadalajara
Huelva
Huesca
Illes Balears
Jaén
La Rioja
Las Palmas
León
Lleida
Lugo
Madrid
Málaga
Melilla
Murcia
Navarra
Ourense
Palencia
Pontevedra
Salamanca
Santa Cruz de Tenerife
Segovia
Sevilla
Soria
Tarragona
Teruel
Toledo
Valencia/València
Valladolid
Zamora
Zaragoza

- Se recorre el listado de juzgados de cada una de las provincias, capturándose, además del municipio, dirección, números de teléfono y código postal, la URL donde figura el detalle de cada uno de ellos:

Directorio de Órganos Judiciales

Selecciona tu provincia y municipio y encuentra el Juzgado que estás buscando:

Provincia	Municipio	Tipo de órgano:
A Coruña	Todos	Todos
<input type="button" value="BUSCAR"/> <input type="button" value="CONSULTAR GOOGLE MAPS"/>		

Municipio	Juzgado	Teléfono/s	Dirección	Código Postal
Arzúa	Sección Civil y de Instrucción del Tribunal de Instancia nº 1	881 99 60 03 881 99 60 04	Rúa de Lugo, 45-47	15810
Betanzos	Sección Civil y de Instrucción del Tribunal de Instancia nº 1	881 88 14 04 881 88 14 05	Praza Irmáns García Naveira, s/n	15300
Betanzos	Sección Civil y de Instrucción del Tribunal de Instancia nº 2	881 88 14 13 881 88 14 14	Praza Irmáns García Naveira, s/n	15300
Betanzos	Sección Civil y de Instrucción del Tribunal de Instancia nº 3	881 88 14 26 881 88 14 27	Praza Irmáns García Naveira, s/n	15300
Betanzos	Sección Civil y de Instrucción del Tribunal de Instancia nº 4	881 88 14 40 881 88 14 39	Praza Irmáns García Naveira, s/n	15300
Carballo	Sección Civil y de Instrucción del Tribunal de Instancia	881 88 11	C/ Río Grande, s/n	15100

Así mismo, se va iterando la acción por cada una de las páginas de listados de esa provincia:

« Previo página 1 de 12 Siguiente »

- Finalmente, para cada uno de los juzgados capturados en cada una de las provincias, se accede al detalle de los mismos, capturándose el listado de jueces titulares, el fax, el email y las coordenadas de geolocalización:

Estás en: Directorio de Órganos Judiciales > A Coruña

Directorio de Órganos Judiciales

Selecciona tu provincia y municipio y encuentra el Juzgado que estás buscando.

Sección Civil y de Instrucción del Tribunal de Instancia nº 2

43°12'26.7"N 8°41'36.6"W
Amplicar el mapa

Dirección: C/ Río Grande, s/n
Código Postal: 15100
Municipio: Carballo
Provincia: A Coruña
Comunidad autónoma: Galicia
Teléfono/s: 881 88 11 90 881 88 11 91
Fax: 881 88 11 92
E-mail: mixto2.carballo@xustiza.gal

5 CONTENIDO

Los campos del dataset son:

Campo	Descripción	Tipo de dato	Ejemplo
Nombre	Nombre del juzgado.	Texto (string)	Juzgado de Primera Instancia nº 3 de Madrid
URL	Dirección web del juzgado.	Texto (string / URL)	https://www.poderjudicial.es/...
Municipio	Municipio en el que se encuentra el juzgado.	Texto (string)	Madrid
Teléfonos	Teléfono/s asignados al juzgado.	Texto (string)	"915123456 915654321"
Dirección	Dirección física del juzgado.	Texto (string)	Calle Capitán Haya, 66
Codigopostal	Código postal del juzgado.	Texto entero (string/int)	28020
DetallesObtenidos	Indica si se han obtenido los datos de detalle del juzgado.	Booleano (true/false)	true
Fax	Número/s de fax del juzgado.	Texto (string)	915987654

Email	Dirección de correo electrónico del juzgado.	Texto (string / email)	juzgado3.madrid@justicia.es
Latitud	Coordenada de latitud del juzgado.	Numérico (float)	40.463667
Longitud	Coordenada de longitud del juzgado.	Numérico (float)	-3.74922
Juez	Nombre de juez del juzgado.	Texto (string)	"María López García"
Provincia	Provincia a la que pertenece el juzgado.	Texto (string)	Madrid

6 PROPIETARIO

El propietario de los datos originales es el Consejo General del Poder Judicial (CGPJ), organismo constitucional español responsable de la supervisión y gestión del poder judicial y sus órganos relacionados.

Los datos extraídos son públicos y se utilizan con fines exclusivamente académicos y de investigación. Por otro lado, se respeta el archivo robots.txt y se implementarán retrasos en las solicitudes HTTP para no sobrecargar el servidor.

No se recopilan datos personales sensibles, únicamente los nombres de los jueces titulares, que son información pública.

Proyectos similares incluyen directorios abiertos de entidades públicas o bases de datos institucionales, como los datasets de ayuntamientos o hospitales.

7 INSPIRACIÓN

El conjunto de datos resulta interesante ya que permite realizar un análisis de la distribución territorial de los órganos judiciales españoles, además de recopilar y facilitar el acceso a datos relevantes para estudios más especializados.

A diferencia de los proyectos similares ya expuestos de ayuntamientos o hospitales, el presente dataset aborda el ámbito judicial, menos representado en los catálogos públicos, por lo que puede resultar útil para análisis de acceso a la justicia o transparencia institucional.

8 LICENCIA

La licencia que hemos escogido es la CC BY-NC-SA 4.0 (Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International)

Hemos escogido esta licencia ya que permite el uso y la distribución del dataset siempre que se haga desde el reconocimiento de autoría. Además, restringe el uso a fines comerciales, asegurándonos de que se utiliza de manera académica.

Por último, la licencia se debe de mantener en trabajos derivados de los datos obtenidos, para que las mismas condiciones que permiten un análisis académico se perpetúen.

9 CÓDIGO

En el código del trabajo está ubicado en este repositorio de GitHub:

 <https://github.com/UOC-Tipologia-y-ciclo-vida-datos/UOC-PRA1-Juzgados-de-España>

En este caso se extraen datos de los juzgados de toda España para posterior análisis.

Cabe reseñar el uso de la librería **BeautifulSoup** para gestionar la captura de datos mediante el uso de consultas XPath y el uso de **Selenium y su integración con el navegador Google Chrome** para capturar el contenido HTML de las páginas desde las que se realiza la captura de los datos.

Nos vimos obligados a utilizar la librería **Selenium** al observar como las peticiones hechas mediante **request** a la página inicial de captura devolvía un error HTTP -código 4xx-, seguramente como mecanismo de protección ante ataques de denegación de servicio o para evitar web scrapping.

10 VÍDEO

 [Enlace](#)

11 TABLA DE CONTRIBUCIONES

Tabla de contribuciones		
	Ángel Barrón Almendros	Pedro Berberana Martín
Investigación previa	Firmado	Firmado
Redacción de las respuestas	Firmado	Firmado
Desarrollo del código	Firmado	Firmado
Participación en el vídeo	Firmado	Firmado

BIBLIOGRAFÍA

- Subirats, L. y Calvo, M. (2018). Web Scraping. Editorial UOC.
- Masip, D. (2019). El lenguaje Python. Editorial UOC.
- Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Capítulo 2: Scraping the Data.
- Munzert, S., Rubba, C., Meiñner, P. y Nyhuis, D. (2015). Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining. John Wiley & Sons.