

Fichas técnicas de las obras contenidas en el Museo del Prado

Fernando Muñoz Martín

Ricardo Santos Patrício

Contexto

Para la realización del trabajo de scrapeo, nos hemos decantado por la página web del Museo del Prado. Para justificar la elección del Museo del Prado como objetivo principal de esta práctica, es importante entender la ola de publicaciones de datos propios que algunos museos empezaron a generar hace años. En este sentido, resulta bastante frustrante que ciertas instituciones no se hayan decidido por dar el paso definitivo a ceder sus datos para el escrutinio de expertos. La situación se da además de una forma bastante asimétrica, mientras que museos de relevancia internacional como el MoMA, los museos Victoria de Australia o el MET se han sumado e incluso contribuido a generar esta corriente, una gran mayoría de museos europeos han decidido no respaldar esta decisión, en su lugar han centrado sus esfuerzos de digitalización en otros campos¹ o se han limitado a publicar sus obras online a modo de catálogo. En este contexto hemos decidido centrarnos en uno de los museos clave del territorio español, el Museo del Prado, el cual decidió mostrar su catálogo online sin publicar sus datos. Resulta especialmente frustrante esta decisión, pues si bien nos permite buscar información puntual de una obra o autor, impide minear su contenido para sacar conocimiento más relevante a nivel histórico-artístico.

Título

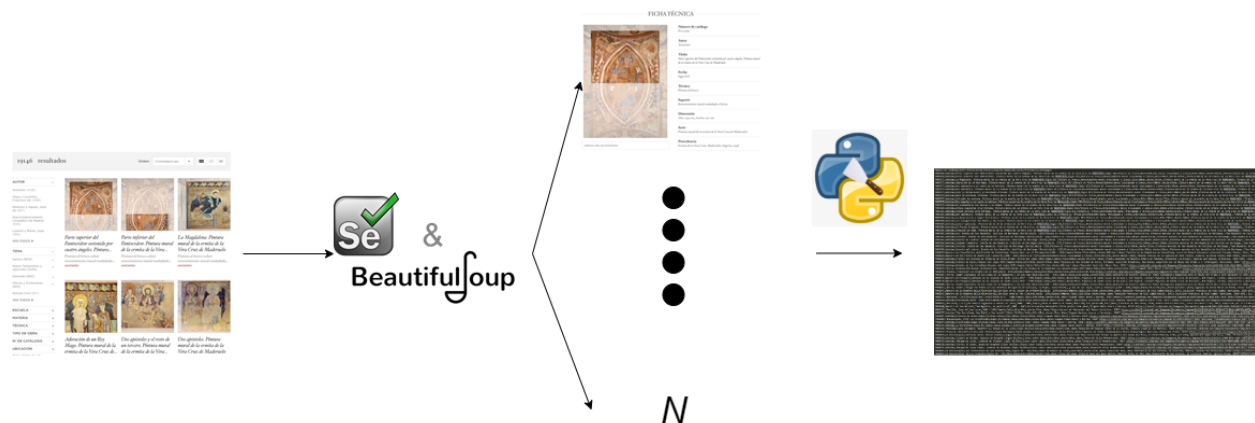
En vista de dar con un título conciso y preciso para los datos contenidos en nuestro dataset, hemos optado por titularlo: Fichas técnicas de las obras contenidas en el Museo del Prado.

¹ Casos como el del museo de El Louvre con el seguimiento de sus visitantes a través de la audioguía <https://senseable.mit.edu/louvre/> (Consultado 30/03/2021) o sus exposiciones virtuales <https://www.louvre.fr/en/online-tours> (Consultado 30/03/2021).

Descripción

El Museo del Prado (y por ende la información contenida en el dataset) se erige como la institución cultural más importante de España. Su importancia no sólo radica en los objetos físicos sino en el legado histórico que representan y que es descriptivo de toda la historia cultural de Europa Occidental. En el dataset resultante de este trabajo encontraremos información técnica sobre las obras de los autores más relevantes de la Historia del Arte, ya sean extranjeros o españoles concretamente desde el siglo XII hasta el siglo XX.

Representación Grafica



Contenido

En el dataset obtenido podemos encontrar información referente a cada una de las obras, donde una instancia supone una obra y cada dimensión se refiere a cada uno de los campos incluidos en las fichas técnicas. Estas características son:

- **Número de catálogo:** identificador único de cada obra dentro del catálogo del Museo del Prado.
- **Autor:** nombre del autor que realizó la obra.
- **Título:** título de la obra.
- **Fecha:** fecha en la que se realizó la obra, comúnmente expresada en siglos.
- **Técnica:** técnica utilizada para realizar la obra.
- **Soporte:** superficie sobre la que se aplicó la pintura en la obra.
- **Dimensión:** columna que contiene la altura y anchura de la obra.

- **Serie:** nombre de la serie a la que pertenece la obra, en caso de pertenecer a alguna.
- **Procedencia:** procedencia de la obra.
- **UrlImagen:** campo donde se incluye la dirección url de las imágenes, con vistas a una posible descarga futura para otros proyectos. Se tiene una carpeta donde se muestra las primeras 15 imágenes descargadas a modo de ejemplo.

Agradecimientos

Al comienzo de este trabajo hacemos una crítica un tanto dura por la ambigüedad de la decisión del Museo del Prado ante la publicación de sus datos de forma abierta, no obstante ahora tenemos que dar la otra cara y agradecer que no podríamos haber recopilado estos datos de no haber sido publicados por el museo en su página web para hacerlos consultables a través de un filtro. Entendemos la dificultad de abrir, exponer o regalar (si se lo quiere llamar así) todos los datos pertenecientes a una colección propia. También aprovechamos para ensalzar otros proyectos digitales que la institución ha creado y que son referencias para las buenas prácticas museísticas como el [MediaLab-Prado](#) o el proyecto [Línea del Tiempo](#).

No obstante, creemos que es necesario aún más, creemos que es necesario poner los datos a disposición de los estudiantes, los académicos e investigadores que aportarían más valor a la colección, la institución y ulteriormente al legado cultural del museo.

No parece lógico que la captura de estos datos la tengamos que llevar a cabo estudiantes² a través de programas de scrapeo sencillos de crear, creemos en su lugar que la línea que todas las instituciones culturales deberían seguir es la de los mencionados MoMA³, Met⁴ o la de los Museos Victoria de Melbourne⁵, más aún en un momento en el que la Historia del Arte está en serio peligro de entregarse

² En <https://github.com/pramodsetlur/Web-Scraper> (Consultado 12/04/2021) el usuario pramodstlur de github (aparentemente creado *ad hoc* para este proyecto) realiza este mismo trabajo sobre la National Portrait Gallery de Washington. A diferencia del proyecto que nos atañe, este fue escrito en java. Se puede leer en los docx que la tarea es un trabajo académico de scrapeo.

³ <https://github.com/MuseumofModernArt/collection> (Consultado 12/04/2021)

⁴ <https://github.com/metmuseum/openaccess> (Consultado 12/04/2021)

⁵ <https://github.com/museumsvictoria/collections-online> (Consultado 12/04/2021)

⁶ No queremos dejar sin mencionar un ejemplo español. La provincia de León ha publicado la información de sus Bienes de Interés cultural online en la página <https://demo.ogooov.com/es/dataset/?id=bienes-interes-cultural> (Consultado 12/04/2021).

al turismo como disciplina y sabiendo que hay tantísimas personas capacitadas deseando utilizar un repositorio abierto para practicar sus habilidades y aportar valor a datasets⁷.

Inspiración

Respondiendo primero al título de este epígrafe, ambos autores percibimos en nuestras conversaciones iniciales que a menudo se da una especie de aura sobre lo cultural, que a todos nos gustaría participar más de lo cultural, pero que su lejanía, hermetismo y muestra a modo de escaparate lo hace muy complicado. Por este motivo decidimos embarcarnos en un proyecto que se centrara sobre uno de los principales museos de nuestro territorio.

Dicho esto, no teníamos ninguna pregunta de investigación en mente antes de comenzar el proyecto, nuestra intención fue en todo momento capturar unos datos relativos a obras de la forma más completa posible. Sí que podemos identificar puntos potencialmente interesantes de investigar para los que el dataset resultante de este trabajo sería un primer paso. A partir de aquí se podrían hacer una investigación descriptiva de la colección del museo, podría tratar el género de los artistas según momento de creación de la obra (a través de los nombres)⁸, se podría analizar la difusión de una técnica concreta por el territorio (especialmente si más museos abrieran sus datos)... Realmente casi cualquier análisis temporal relativo a cada una de las dimensiones incluidas sería factible e interesante. También hemos tenido en cuenta incluir enlaces a las imágenes por si eventualmente pudieran ser utilizadas para alimentar redes neuronales con fines investigadores o incluso artísticos a través de GAN's o Autoencoders⁹.

En lo relativo a la Historia del Arte, en muchos casos no se necesitan algoritmos complejos para extraer información, la simple capacidad de mostrar la información de manera adecuada permite responder muchas preguntas relacionadas con los patrones de comportamiento de los artistas así como relativos a la creación de sus obras.

⁷ Shin, Y., *Using MoMA's collection dataset to visualize whose stories are missing*
<https://towardsdatascience.com/using-momas-collection-dataset-to-visualize-whose-stories-are-missing-76a8960a33c2> (Consultado 12/04/2021)

Utilizamos este ejemplo a modo ilustrativo, la cantidad de tutoriales que el dataset del MoMA como referencia es absolutamente inabarcable.

⁸ Justo en los últimos meses hemos visto cómo el Prado está recuperando el trabajo de Artemisia Gentileschi para situarla en el lugar histórico que se merece.
https://elpais.com/cultura/2020/06/19/babelia/1592589223_823778.html (Consultado 12/04/2021).

⁹ A modo de muestra pueden encontrarse en el repositorio las 15 primeras imágenes.

Licencia

Después del discurso establecido para este trabajo, no sería coherente elegir una licencia que no fuera la Creative Commons 1.0, mediante la cual no pretendemos tener ningún derecho autoral sobre la obra resultante de este trabajo.

Sí que queremos remarcar que su creación es parte de un proceso de scraping que, en la medida de nuestro conocimiento, ha respetado las indicaciones incluidas en el archivo robots.txt¹⁰ a día 12 de abril de 2021.

Firmas

Contribuciones	Firma
Investigación previa	Fernando Muñoz y Ricardo Santos.
Redacción de las respuestas	Fernando Muñoz y Ricardo Santos.
Desarrollo código	Fernando Muñoz y Ricardo Santos.

¹⁰ <https://www.museodelprado.es/robots.txt> (Consultado 12/04/2021)