

# TIPOLOGÍA Y CICLO DE VIDA DE LOS DATOS

## PRÁCTICA 1: WEB SCRAPING

Leticia García Ramos

Javier Batuecas Rodríguez

12 de noviembre de 2018

## Contenido

CATÁLOGO DE MESSIER .....	3
Autores .....	3
Importancia de los datos. Contexto e imagen .....	3
Contenido .....	4
Agradecimientos .....	4
Información sobre los autores .....	4
Inspiración .....	4
Licencia .....	5
Ficheros incluidos .....	5
Referencias utilizadas para la realización de la práctica .....	5

## CATÁLOGO DE MESSIER

En esta práctica se recoge en un csv la información más relevante de los objetos (cúmulos, galaxias y nebulosas) incluidos en el catálogo de Messier.

### Autores

Javier Batuecas Rodríguez

Leticia García Ramos

### Importancia de los datos. Contexto e imagen

Durante los años de 1758 a 1782, Charles Messier, un astrónomo francés (1730 - 1817), compiló una lista de aproximadamente 100 objetos difusos que eran difíciles de distinguir de los cometas a través de los telescopios de la época.

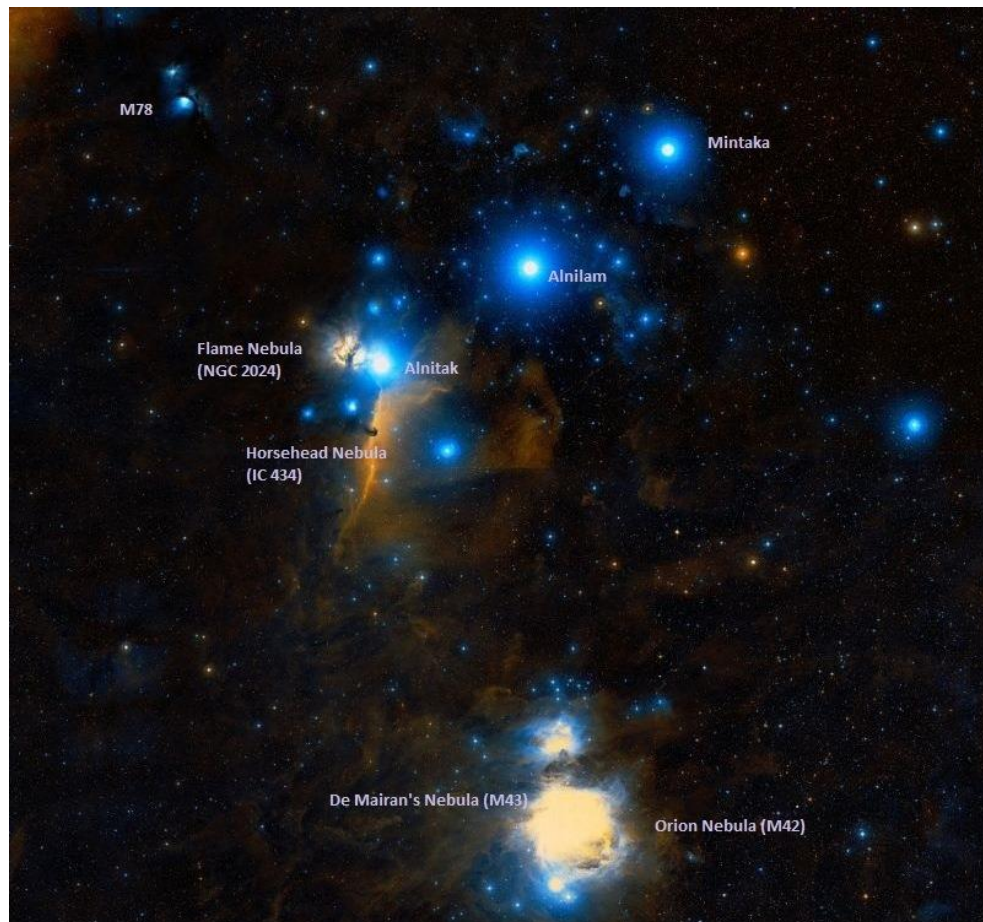


Ilustración 1 Imagen obtenida de <http://www.messier-objects.com/>

## Contenido

Incluye campos de nombre del objeto, información básica e información característica del mismo como puede ser el brillo. Además, se incluye enlace a la web para obtener una descripción más detallada. En caso de considerarse interesante o necesario, se podría extraer esta información de detalle en el propio dataset

## Agradecimientos

Durante la realización de esta práctica se han consultado diferentes webs con información sobre el catálogo de Messier:

<http://www.messier-objects.com/>

<http://www.catalogomessier.com/>

<https://www.astronomia-iniciacion.com/el-catalogo-messier.html>

Pero la recopilación de estos datos ha sido posible gracias a la recopilación de la información de los objetos de Messier por Guy McArthur, Mark Elowitz, Hartmut Frommert y Christine Kronberg autores de la BBDD de <http://www.messier.seds.org/>, página de la que se ha extraído la información.

## Información sobre los autores

A continuación se indican una serie de estudios publicados:

- Elowitz, R. Mark, Scholarly Research Project, Space Studies Department, University of North Dakota (2012), A Search for Variations of the Fine Structure Constant Using Quasar Absorption Lines.
- Dayton, David C., Elowitz, Robert M., Lasche, James B., and Hanes, Stephen A., Simulating Range-Amplitude Returns for the LARRA (Laser Radar for Recognition and Assessment) Program, Proc. SPIE Vol. 4490, p. 256-265., 2001.

## Inspiración

Como astrónomos aficionados, consideramos muy útil una recopilación de los objetos celestes más importantes y/o interesantes para observar con equipos de calidad no profesional. Para astrónomos profesionales, como hemos mencionado antes, la contribución de Messier es la base de muchos descubrimientos posteriores.

## Licencia

Hemos seleccionado para este trabajo: Released Under CC BY-NC-SA 4.0 License (condiciones recogidas en el fichero LICENSE.txt) Licencia de dominio público pero que no permite el uso de datos con ningún fin comercial. Nos hemos decantado por esta licencia ya que consideramos que los datos pueden ser usados como ayuda para diversos estudios y así se podrá acceder libremente a ellos, pero no podrán usarse con fines comerciales.

## Ficheros incluidos

- README.md: Breve descripción del trabajo.
- messier\_catalogue.csv: fichero que recoge la información que se ha considerado relevante y, por tanto, se ha extraído de <http://www.messier.seds.org/>.
- messierCatalogueScraper.py: contiene la implementación del código que genera el conjunto de datos a partir de la base de datos de la web indicada anteriormente.
- LICENSE.txt: archivo que incluye la licencia bajo la que se ha calificado el *dataset*.
- PRACTICA1.pdf: Respuestas a las preguntas (texto incluido en esta Wiki).

## Referencias utilizadas para la realización de la práctica

- <https://jarroba.com/scraping-python-beautifulsoup-ejemplos/>
- <https://docs.python.org/3/tutorial/index.html>
- <https://www.lawebdelprogramador.com/foros/Python/1574263-beautifulSoup-y-extraccion-de-datos-de-tablas.html>