

Pràctica 1 – Tipologia i cicle de vida de les dades

Assignatura: M2.951 / Semestre: 2022/23-2 / Data: 10-04-2023

URL del lloc web triat: <https://www.vesselfinder.com/>

Autors:

- Francisco J. Bastida López – fbastidal@uoc.edu
- Ivan Benaiges Trenchs – ibenaiges@uoc.edu

Resolució dels apartats

1. Context

El sistema d'identificació automàtica (AIS) de vaixells és un sistema de transmissió per VHF de posició, velocitat i estat de navegació concebut per raons de seguretat perquè altres vaixells o estacions terrestres puguin fer un seguiment de les embarcacions i en darrera instància poder salvar vides en el mar. L'Organització Marítima Internacional (OMI) obliga al seu ús a les embarcacions de certes característiques (1). Pel propi disseny del sistema, les dades s'emeten per radiofreqüència en obert a fi que qualsevol altre transceptor (altres vaixells pròxims) pugui reenviar la senyal. Això ha derivat en un ús secundari no planificat i emergent de fer aquestes dades visibles per internet mitjançant múltiples proveïdors de serveis.

Les dades en obert que es poden descarregar per internet, en general són històriques i agregades; mentre que les dades actuals es poden consultar en pàgines web de diferents proveïdors susceptibles de fer web scraping. Entre tots els proveïdors que publiquen aquestes dades hem triat el lloc web de *Vessel Finder*, perquè complementa les dades de posicionament AIS amb una base de dades amb les característiques de cada vaixell. La pàgina principal a partir de la qual s'ha programat el web scraping és: <https://www.vesselfinder.com/es/vessels>

2. Títol

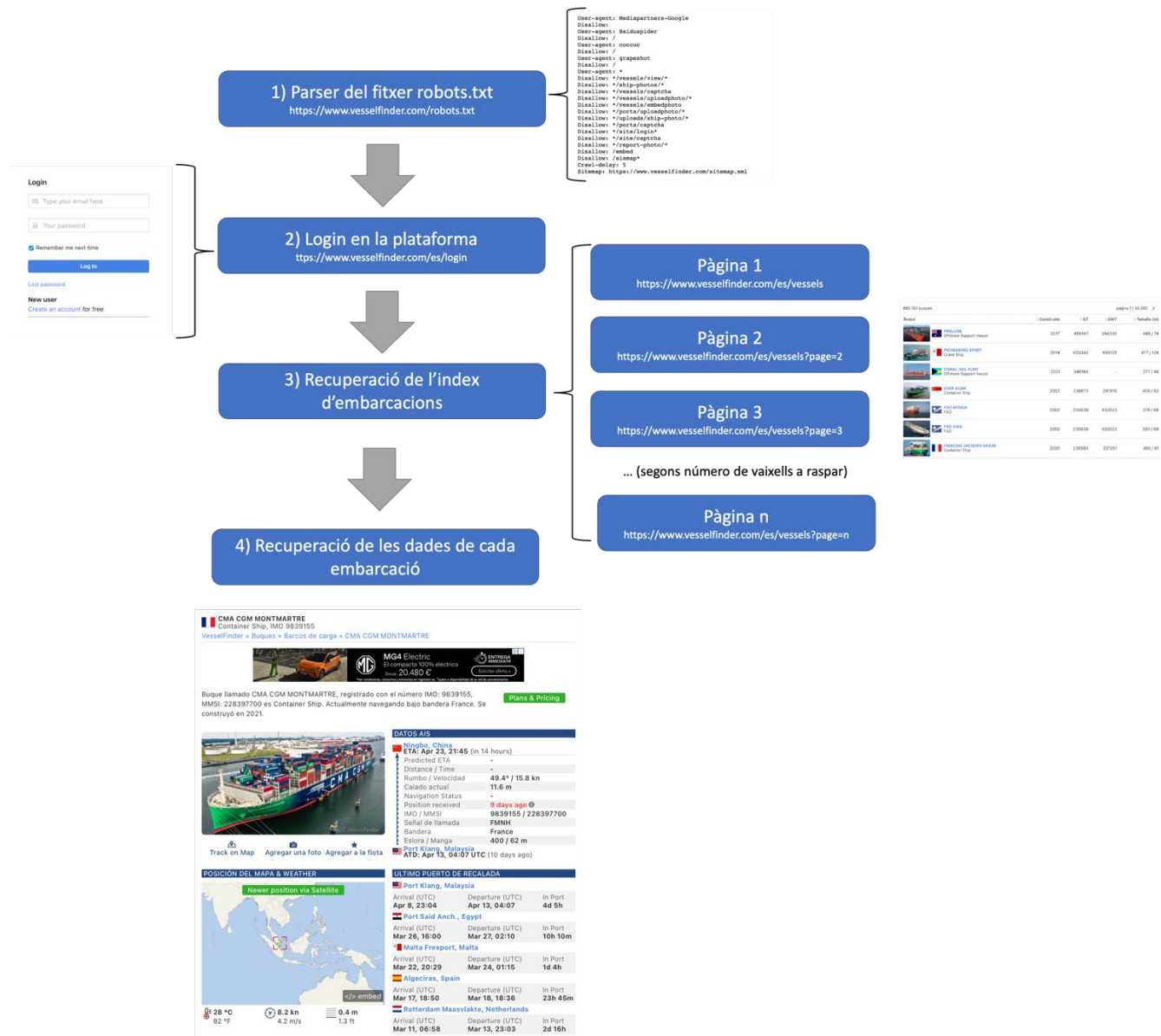
Navegació mundial d'embarcacions

3. Descripció del dataset

El dataset conté les dades sobre les grans embarcacions i la seves característiques de navegació disponibles en el moment de fer el scraping sobre la darrera lectura AIS enregistrada amb dades sobre els cinc darrers ports visitats.

4. Representació gràfica

Mostram a continuació un gràfic amb les pàgines de dades rellevants del projecte triat:



5. Contingut

El dataset té una observació (fila) per a cada vaixell amb un conjunt de variables (columnes) amb les característiques de l'embarcació:

- **nombre**: Nom del vaixell
- **tipo**: Tipus d'embarcació
- **eslora**: Eslora en metres
- **manga**: Màniga en metres
- **calado**: Calat en metres
- **año**: Any de construcció del vaixell
- **origen**: Lloc de construcció
- **bandera**: Nacionalitat de l'embarcació
- **imo**: Identificador IMO

Adicionalment hi ha una sèrie de variables sobre dades de la seva darrera lectura de dades enregistrada:

- **puertoDestino**: Port de destinació
- **urlPuertoDestino**: Pàgina web del port de destinació
- **puertoOrigen**: Port d'origen de la seva travesia
- **urlPuertoOrigen**: Pàgina web del port d'origen
- **eta**: Data i hora estimada d'arriba a destí
- **distancia**: Milles de distància a destí
- **tiempo**: temps previst d'arribada
- **rumbo**: rumbo actual del vaixell
- **velocidad**: nudos actuals
- **caladoActual**: calat actual
- **estadoNavegacion**: Estat de la navegació actual
- **ultimaPosicionRecibida**: Data i hora de la darrera posició rebuda
- **MMSI**: Identificació de la llicència de radio
- **signal**: senyal de la cridada
- **GT**: Tonelatge brut del vaixell (correspon al volum total intern de l'embarcació)
- **DWT**: Tonelatge de pes mort (mesura del pes total que pot transportar el vaixell)
- **TEU**: Capacitat de càrrega en número de contenidors estàndards que pot dur
- **crudo**: Número de barrils de petroli que pot dur en el cas de petroliers.
- **grano**: Mesura del volum dels depòsits de grans
- **fardo**: Número de fardos que pot transportar

Per últim, també es recull dades dels cinc darrers ports que ha visitat l'embarcació:

- **puerto0n_nombre**: Nom del port
- **puerto0n_llegada**: Data i hora d'arribada al port
- **puerto0n_salida**: Data i hora de sortida del port
- **puerto0n_tiempo**: Temps d'estada en el port

Cal dir que algunes de les dades no estan disponibles per comptes gratuïts com és el cas del compte emprat en la pràctica; no obstant, el programa recull totes les dades; només cal entrar amb un usuario premium per poder obtenir-les totes.

6. Propietari

Tal i com s'ha explicat en el punt 1, les dades AIS es publiquen en obert per radiofreqüència i en aquest medi són lliures. El lloc web *VesselFinder.com* recull aquestes dades per radiofreqüència i les publica per internet amb dades complementàries sobre característiques de les embarcacions. Des d'aquest moment, les dades són propietat de *VesselFinder.com*, la qual les cedeix d'acord amb els seus termes d'ús (4) sempre que no es venguin, lloguin o sublllicencien i es mencioni a *VesselFinder.com* com a font.

7. Inspiració

Tot i que el sistema AIS és va dissenyar inicialment per evitar col·lisions entre grans vaixells, actualment la disponibilitat en obert de les dades ha permès el desenvolupament de moltes altres aplicacions com les que citen en (3): monitorització i control de la flota pesquera, seguretat marítima per controlar activitats dins la zona econòmica exclusiva de cada país, ajudes a la navegació (corrents, condicions climàtiques...), cerca i salvament marítim, investigació d'accidents, protecció d'infraestructures marines, seguiment de flotes i carregament i per últim, estadístiques i anàlisi de dades.

L'ús d'aquestes dades per al seu anàlisi ha permès tenir dades en temps real dels fluxos comercials, estimar indicadors econòmics més ràpids, tràfic dels ports, nivell d'emissions de gasos d'efecte hivernacle i moltes altres estadístiques interessants. Tant és l'interès sobre aquestes dades que les Nacions Unides organitzen cada any un esdeveniment anomenat "Setmana de la dada AIS". A la pàgina web de l'esdeveniment (2) es pot veure el ventall de ponències amb preguntes a les que poden donar resposta aquestes dades, entre les que podem ressaltar:

Aquesta interès mundial per aquestes dades ha estat la motivació principal per enfocar aquesta pràctica.

8. Llicència

La llicència de publicació de les dades ha de ser CC BY-NC-SA 4.0 License perquè els termes d'ús de la font de dades ja són així:

- S'ha de citar la font
- No es pot fer un ús comercial de les dades
- Qualsevol adaptació s'ha de compartir en les mateixes condicions

9. Codi

El codi implementat per obtenir el dataset s'ha desenvolupat en Python fent servir la tecnologia Selenium i es pot descarregar des de Github a <https://github.com/fbastidal/Practica-1>

Els aspectes més importants a ressaltar del codi són:

- S'ha programat en Python amb l'ajuda de la llibreria de Selenium.
- El codi està estructurat en dos fitxers:
 - main.py amb els paràmetres de l'aplicació, el menú principal i les cridades a les funcions d'scrapper
 - scraper.py: Amb la definició de les classes i les funcions principal d'scrapper
- Funcionalitat:
 - En primer lloc es llegeix i es fa un parser del fitxer robots.txt per respectar les restriccions i la configuració del mateix (especialment el paràmetre *craw-delay*)
 - A continuació es demana el número d'embarcacions que es vol carregar (actualment la plataforma té de l'ordre de 680.000 embarcacions) i el nom del fitxer de sortida del resultat.
 - A continuació es fa un login a la plataforma amb la finalitat de el raspat de dades es faci amb els privilegis del compte (si és un compte premium es pot accedir a més dades).
 - A continuació es procedeix a recollir la URL de tant de vaixells com es volen raspar; per aixó es va navegant pel cercador tantes pàgines com sigui necessari fins a arribar al número d'embarcacions desitjades o quan s'acabi la llista.
 - Una vegada que es disposa de la llista de URL de tots els vaixells, es va un a un recollint totes les dades individuals
 - Acabada la llista d'embarcacions es procedeix a enregistrar les dades en el fitxer.

10. Dataset

El dataset obtingut s'ha publicat en format CSV a Zenobo: <https://doi.org/10.5281/zenodo.7856489>

11. Vídeo

El vídeo explicatiu de la pràctica està a: https://drive.google.com/file/d/1TMU1ZIB34IOh7JD-MI_Q9x48dOYZq6kU/view?usp=share_link

12. Bibliografia

- **(1) Resolució A.1106(29) adoptada el 2 de desembre de 2015 per l'Organització Marítima Internacional:**
[https://wwwcdn.imo.org/localresources/en/OurWork/Safety/Documents/AIS/Resolution%20A.1106\(29\).pdf](https://wwwcdn.imo.org/localresources/en/OurWork/Safety/Documents/AIS/Resolution%20A.1106(29).pdf)
- **(2) Porta del esdeveniment sobre dades AIS organitzat per les Nacions Unides:**
<https://unstats.un.org/unsd/trade/events/2020/AisDataWeek/default.asp>
- **(3) Automatic identification System: Cerca en wikipedia:**
https://en.wikipedia.org/wiki/Automatic_identification_system
- **(4) Termes d'ús de les dades de VesselFinder.com:** <https://www.vesselfinder.com/es/terms>

Contribucions	Signatura
Investigació prèvia	Francisco Bastida López, Ivan Benaiges Trenchs
Redacció de les respostes	Francisco Bastida López, Ivan Benaiges Trenchs
Desenvolupament del codi	Francisco Bastida López, Ivan Benaiges Trenchs
Participació en el vídeo	Francisco Bastida López, Ivan Benaiges Trenchs