

Arnau Arasa, Cinta

# Curs 2022-23

**Citizen Science Projects Recommendation System Based On The Spanish Elementary School Curriculum**



Directores: Patricia Santos i Miriam Calvera

GRAU EN ENGINYERIA MATEMÀTICA EN CIÈNCIA DE DADES

Treball de Fi de Grau

# GRAU EN ENGINYERIA EN xxxxxxxxxxxx

*[Pàgina en blanc]*

*[S’han marcat els apartats del treball que haurien d’anar en pàgina senar o dreta de la publicació (situació de pàgines que ajuda a estructurar i a donar prioritat formal als diferents apartats).*

*Hi ha apartats on no s’indica res. En aquests cas l’estudiant pot triar la situació, en pàgina parell o senar, segons li convingui per al compaginat final.*

*Es recomana que les pàgines blanques no portin número de foli (es compten, però el número no s’imprimeix)]*

*Si s’imprimeix a simple cara no s’han de deixar pàgines en blanc*

Dedicatòria [opcional] [12 punts]

*[Pàgina en blanc]*

**Acknowledgements**

Text dels agraïments [12 punts]

*[Pàgina en blanc]*

**Abstract**

Resum del treball en la llengua que es presenti [12 punts]

Extensió màxima recomanada: 150 paraules

**Resumen [en una 2a llengua. Ex. Resumen]**

Resum del treball en una llengua diferent a la utilitzada [12 punts]

Extensió màxima recomanada: 150 paraules

**Resum [en una 3a llengua. Ex. Abstract]**

Resum del treball en una llengua diferent a la utilitzada [12 punts]

Extensió màxima recomanada: 150 paraules

*[Aquest apartat,* ***sempre*** *hauria de començar en pàgina senar, pàgina dreta de la publicació]*

*[Pàgina en blanc]*

**Table of contents**

*[Taula de contingut del treball, parts en què està dividida] [12 punts]*

*[Aquest apartat també ha de començar en pàgina senar]*

**Table of figures [opcional]**

**Llista de taules [opcional]**

**1. INTRODUCTION**

**1.1 Títol de l’apartat [arial 14 punts]**

Text de l’apartat [12 punts]

a) Títol de la subdivisió [14 punts]

Text de la subdivisió a) [12 punts]

b) Títol de la subdivisió [14 punts]

Text de la subdivisió b). Nota a peu de pàgina [[1]](#footnote-1) [12 punts]

*[Cada capítol ha de començar en pàgina senar]*

**2. STATE OF THE ART**

2.1 Títol de l’apartat [14 punts]

Text de l’apartat [12 punts]

a) Títol de la subdivisió [14 punts]

Text de la subdivisió a) [12 punts]

b) Títol de la subdivisió [14 punts]

Text de la subdivisió b). Nota a peu de pàgina [[2]](#footnote-2) [12 punts]

*[Cada capítol ha de començar en pàgina senar]*

**3. DESIGN AND IMPLEMENTATION**

3.1 Environment set-up and selection of tools

3.1.1) Tools Used

In order to perform the analysis of the two platforms of Citizen Science projects, various tools will be used throughout the entirety of the thesis process. The following descriptions of these tools aim to clarify their significance and demonstrate their application within the project.

**3.1.1.1) Python**

Python[[3]](#footnote-3) is a high-level programming language that is widely used for general-purpose programming. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming. Python’s versatility and ease of use have made it a go-to language for a wide range of applications, including web development, data analysis, machine learning, artificial intelligence, scientific computing, and automation [4]. For this reason, and given the fact that Python has been the most used programming language throughout the degree, it has been the chosen language for the Citizen Science projects’ extraction and the building of the recommendation system.

The libraries that have been used in the scripts (both the project extraction and the recommendation system) are the following:

* Pandas[[4]](#footnote-4): “It is a powerful open-source Python library widely used for data manipulation, analysis, and exploration”. It provides highly efficient data structures and data analysis tools, making it an essential tool for working with structured data, creating DataFrame objects to manipulate data with integrated indexing, and much more functionalities to work with data sets [2].
* Request[[5]](#footnote-5): “It is a popular tool used for making HTTP requests and interacting with web services. It provides a convenient and user-friendly interface to send HTTP requests, handle responses, and perform various operations related to web communication” [3].
* BeautifulSoup[[6]](#footnote-6): “Python library used for web scraping and parsing HTML or XML documents” [1].
* (…)

**3.1.1.2) Jupyter Notebook**

“Jupyter Notebook[[7]](#footnote-7) is an open-source web-based interactive computing environment widely used for data analysis, visualization, and prototyping” [6]. It supports many programming languages (including Python) and allows for interactive data analysis and visualization, enhancing the exploratory data analysis process.

**3.1.1.3) BeautifulSoup**

“BeautifulSoup is a popular Python library used for web scraping and parsing HTML or XML documents” [1]. It provides a convenient way to extract and navigate data from web pages by simplifying the process of locating and manipulating elements within the document structure. BeautifulSoup can be combined with other Python libraries, such as requests, to fetch web page content and then parse and extract the desired information.

**3.1.1.4) GitHub**

“GitHub[[8]](#footnote-8) is a web-based platform for version control and collaborative software development”. It provides a centralized hub where developers can store, manage, and collaborate on their code repositories [5]. A GitHub repository[[9]](#footnote-9) has been created to store and manage all the scripts with the corresponding code of this thesis.

3.1.2) Jupyter Notebook Scripts

In order to extract all the information about the Spanish and Catalan Citizen Science projects, the script “projects\_extraction.ipynb” has been generated. All the information is extracted using the Python library BeautifulSoup as it will be later explained.

The other script named “recommendation\_system.ipynb” (…).

These Jupyter notebook scripts can be found in the Github repository so that anyone interested in Citizen Science or any teacher wanting to use the recommendation system can consult them (see AppendixX to consult the locations of the scripts).

*[Cada capítol ha de començar en pàgina senar]*

**BIBLIOGRAPHY**

1. BeautifulSoup Documentation. (n.d.). Retrieved May 20, 2021, from <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>.
2. Pandas Documentation. (n.d.) Retrieved May 20, 2021, from <https://pandas.pydata.org/docs/>.
3. Requests Documentation. (n.d.). Retrieved May 20, 2023, from <https://docs.python-requests.org/en/latest/>.
4. Python Software Foundation. (n.d.). Python. Retrieved May 20, 2023, from <https://www.python.org/>.
5. GitHub. (n.d.). About GitHub. Retrieved May 20, 2023. from <https://github.com/about>.
6. Project Jupyter. (n.d.). Jupyter Notebook. Retrieved May 20, 2023, from <https://jupyter.org/>.

*Cal documentar les fonts bibliogràfiques utilitzades amb un format de citació estàndard:*

[*http://guiesbibtic.upf.edu/tic/tfg*](http://guiesbibtic.upf.edu/tic/tfg) *(apartat* ***Presentació del treball*)**

*La Biblioteca de la UPF ofereix el gestor de bibliografies Mendeley, que us permet crear la vostra base de dades personal de referències bibliogràfiques en línia, importar referències automàticament des de diferents recursos d'informació, extreure llistes i generar bibliografies en diferents estils de citació, i incorporar les citacions i llistes de bibliografia als vostres documents de text*

[*Gestor de bibliografies Mendeley*](http://guiesbibtic.upf.edu/mendeley)

*Exemple de cita bibliogràfica [12 punts]*

Heery, M. “Organització de la biblioteca: Repàs d'estructures.” *Item: Revista de biblioteconomia i documentació 23.2 (1998): 8-15.*

Text de la cita bibliogràfica [12 punts]

*[Aquest apartat ha de començar en pàgina senar]*

1. Text de la nota al peu de pàgina [10 punts] [↑](#footnote-ref-1)
2. Text de la nota al peu de pàgina [10 punts] [↑](#footnote-ref-2)
3. Python Website (2023, May 20) <https://www.python.org/>. [↑](#footnote-ref-3)
4. Pandas Documentation (2023, May 20) <https://pandas.pydata.org/docs/>. [↑](#footnote-ref-4)
5. Request Documentation (2023, May 20) <https://docs.python-requests.org/en/latest/>. [↑](#footnote-ref-5)
6. BeautifulSoup Documentation (2023, May 20) <https://www.crummy.com/software/BeautifulSoup>. [↑](#footnote-ref-6)
7. Jupyter Notebook Website (2023, May 20) <https://jupyter.org/>. [↑](#footnote-ref-7)
8. GitHub Website (2023, May 20) <https://docs.github.com/en>. [↑](#footnote-ref-8)
9. GitHub repository <https://github.com/Cintaa1223/TFG>. [↑](#footnote-ref-9)