
Web scraping

PID_00256967

Laia Subirats Maté
Mireia Calvo González

**Laia Subirats Maté**

Enginyera de Telecomunicacions per la Universitat Pompeu Fabra (2008), màster en Telemàtica per la Universitat Politècnica de Catalunya (2009) i doctora en Informàtica per la Universitat Autònoma de Barcelona (2015). Des de 2009, treballa com a investigadora a Eurecat (Centre Tecnològic de Catalunya) aplicant la ciència de dades a diferents camps com ara la salut, el medi ambient o l'educació. Des de 2016, col·labora amb la UOC com a docent en el màster de Data Science i en el grau d'Informàtica. És especialista en intel·ligència artificial, ciència de dades, salut digital i representació del coneixement.

**Mireia Calvo González**

Enginyera de Telecomunicacions per la Universitat Politècnica de Catalunya (2011), màster en Enginyeria Biomèdica per la Universitat de Barcelona i la Universitat Politècnica de Catalunya (2014) i doctora en Processament de Senyals i Telecomunicacions per la Universitat de Rennes 1 i en Enginyeria Biomèdica per la Universitat Politècnica de Catalunya (2017). Des de 2012, ha treballat com a investigadora en diferents entorns acadèmics, clínics i industrials, aplicant el processament de dades a l'estudi de diferents malalties cardíaques i respiratòries. Des de 2017, col·labora amb la UOC com a docent en el màster de Data Science.

L'encàrrec i la creació d'aquest recurs d'aprenentatge UOC han estat coordinats per la professora: Isabel Guitart Hormigo (2019)

Primera edició: febrer 2019
© Mireia Calvo González i Laia Subirats Maté
Tots els drets reservats
© d'aquesta edició, FUOC, 2019
Av. del Tibidabo, 39-43, 08035 Barcelona
Disseny: Manel Andreu
Realització editorial: Oberta UOC Publishing, SL

Cap part d'aquesta publicació, incloent-hi el disseny general i la coberta, no pot ser copiada, reproduïda, emmagatzemada o transmesa de cap manera ni per cap mitjà, tant si és elèctric com químic, mecànic, òptic, de gravació, de fotocòpia o per altres mètodes, sense l'autorització prèvia per escrit dels titulars del copyright.

Continguts

Mòdul didàctic 1

Web scraping

Laia Subirats Maté, Mireia Calvo González i Isabel Guitart Hormigo

1. Per què i com fer *web scraping*?
2. Primers passos per a fer *web scraping*
3. *Web scraping* de contingut gràfic i audiovisual
4. Emmagatzematge i compartició de dades
5. Prevenició del *web scraping*
6. Resolució d'obstacles en *web scraping*
7. Aspectes legals
8. Millors pràctiques i consells
9. Exemples de *web scraping* i casos d'èxit

