

Estudi de la presència de conceptes relatius a la transformació digital en l'àmbit de les contractacions públiques europees

David Roche Valles i Silvia Galan Martínez

1. Context

La nostre pràctica de "Web Scraping" sorgeix dins d'un projecte basat en l'interès en conèixer la distribució geogràfica de les contractacions públiques relacionades amb la transformació digital o contractacions "digitals" dins de l'àmbit europeu.

Les contractacions públiques europees provenen de diverses institucions estatals, comarcals o municipals de tot l'espai europeu. La Unió Europea (UE) posa a disposició de la ciutadania una plataforma per gestionar tots els aspectes relatius a la contractació pública europea, el Tenders Electronic Daily. Aquesta plataforma, entre les seves funcionalitats, permet accedir a les dades de tots els contractes dels últims anys on apareix informació relativa a la data de publicació, la institució que presenta el projecte, el propi projecte amb la descripció i altres conceptes. De fet, és fàcil obtenir totes les dades dels projectes oferts en formats plans o fulls de càlcul i també disposa d'APIs per accedir a la informació.

L'objectiu de context és crear clústers o grups de zones europees basant-se en si els projectes oferts són relatius a la transformació digital, en concret amb les tecnologies de la informació (IT). Per això, es necessita definir, primer de tot, les entitats geogràfiques a nivell europeu. Degut a que els diferents països presenten diferents formes de dividir el territori, la UE va crear una classificació per regions anomenada NUTS que classifica les localitzacions geogràfiques a 3 nivells anomenats NUTS1, NUTS2 i NUTS3. Aquesta classificació s'acostuma a utilitzar quan es necessita geo-localitzar a nivell europeu en diferents escenaris, com per exemple, en recerca, projectes, estudis i economia. El nivell NUTS3 és l'escollit i correspondria a nivell espanyol a les comarques o els municipis.

Per tant, el projecte de context, pretén descriure les contractacions europees des del punt de vista de la relació amb la transformació digital a nivell NUTS3 amb la intenció de crear un mapa descriptiu d'agrupacions o clústers de NUTS3 semblants en quant a contractació pública referent a projectes relacionats amb la transformació digital.

Un dels problemes més importants d'aquest objectiu és que les dades sobre les contractacions públiques contenen el camp NUTS3 però, sorprenentment, no està informat. Donat que aquest camp és vital per poder realitzar l'estudi de context, la compleció d'aquesta dada per cadascun dels projectes és essencial per dur a terme l'estudi. Inclús, a partir d'altres camps de la base de dades (o taula) com la ciutat, el poble o el codi postal és difícil arribar a trobar el NUTS3 corresponent ja que no hi ha una relació directa o els camps que es vincularien també estan pràcticament buits o amb molts valors perduts. En aquest punt és on intervé el "Web scraping".

Cada contractació de la qual informa la base de dades, conté, també, un camp amb una url a un document en format pdf on surt tota la informació referent al projecte. En aquests pdf's si que està

informada la variable NUTS3. Ara bé, cal accedir-hi als pdfs i extreure el camp del NUTS3 per cadascun dels 200.000 projectes anuals.

És aquí on es perfila l'objectiu d'aquesta pràctica de "Web scraping". Completar la informació de la taula de les contractacions per informar la variable NUTS3 i així poder encetar l'objectiu del projecte de context i trobar la distribució per NUTS3 de l'etiqueta "digital" en les contractacions públiques europees. Respecte el concepte "contractació relacionada amb la transformació digital" hem utilitzat un camp que es diu CPV, que classifica segons el seu objectiu cadascuna de les contractacions.

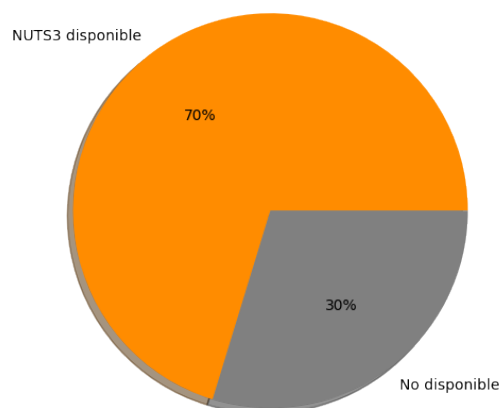
2. Descripció del dataset

Tal com s'explica a l'apartat del context, la base de dades utilitzada en aquest estudi són les dades de contractacions públiques europees de l'any 2018 publicades a la plataforma TED, Tenders Electronic Daily.

Hi ha informació sobre 227.022 contractacions públiques europees diferents identificades per la variable "ID_NOTICE_CN". La taula inicial de dades, s'ha restringit per temes de mida i simplificació a dades relatives a Espanya on hi ha 13.058 contractacions públiques de les quals 93 estan cancel·lades (variable CANCELLED =1). Per tant, la base de dades final usada com a input, conté 12.956 contractacions diferents a nivell espanyol. Posteriorment, per tal d'assolir el nostre objectiu, cal netejar el dataset de forma que tinguem ben informades les dades que contenen la informació del NUTS3, això ens permet mantenir la taula final on hi ha 9.111 registres que informen sobre contractacions espanyoles d'institucions públiques no cancel·lades i amb el NUTS3 ben informat.

Les variables de la taula de dades inicial, es poden trobar descrites aquí. La taula de dades inicial conté 64 variables de les quals s'han seleccionat les que tenen interès de cara a l'objectiu de l'estudi, per suposat NUTS3 i CPV però també d'altres que poden completar la descripció del fenomen d'interès.

En el següent gràfic de sectors, podem observar el percentatge de projectes a Espanya dels quals s'ha obtingut informació completa sobre el NUTS3, després d'aplicar el codi *gets_nuts.py*, amb comparació dels que no.

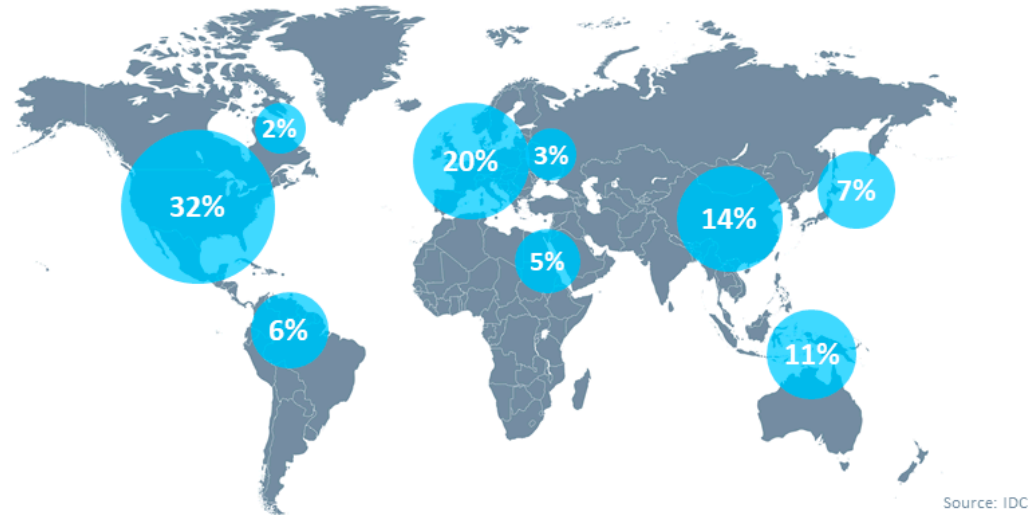


3. Representació gràfica

La següent imatge mostra la idea del projecte, obtenir la distribució, per zones estadístiques, de la contractacions relatives a la transformació digital o tecnologies de la informació.

The Global Information Technology Industry: \$5.2 Trillion

Estimated 2020 spending at constant currency | Encompasses hardware, software, services and telecommunications



CompTIA. (2020). *IT industry outlook 2020*. Recuperat de <https://www.comptia.org/content/research/it-industry-trends-analysis>

4. Contingut

Descripció de les dades

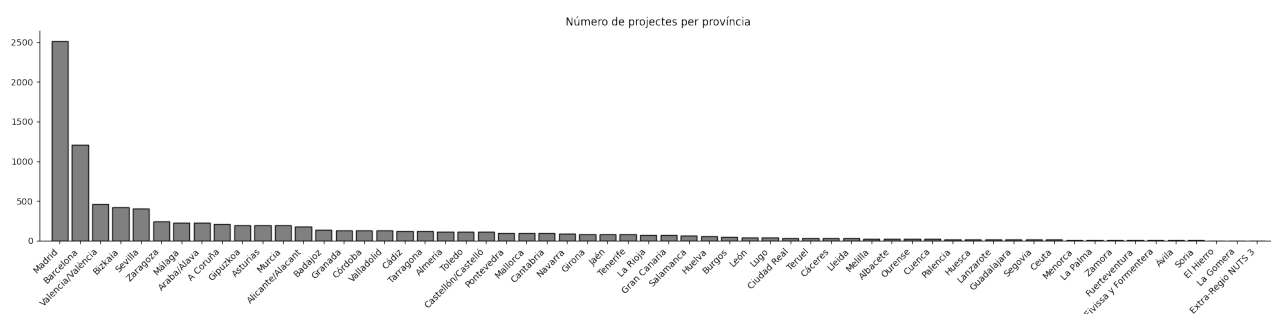
La taula de dades final conté 9.111 contractacions públiques europees d'Espanya, informades en 13 variables. Aquestes variables detallen aspectes relacionats amb la pròpia entitat o organització de contractació, la ubicació i aspectes relatius a la pròpia contractació, com el nom o objectiu, la durada, el valor econòmic en euros, el tipus, i el codi d'objectiu, a més de la variable per tal de saber si la contractació està relacionada amb IT o no. Per tant, a partir d'aquesta informació, podem estudiar la distribució de les contractacions per les diferents regions espanyoles segons si són o no relatives a la transformació digital (a partir de l'atribut CPV), permetent l'anàlisi amb altres indicadors com l'econòmic o la durada.

A continuació es mostra detalladament les variables seleccionades per la taula definitiva:

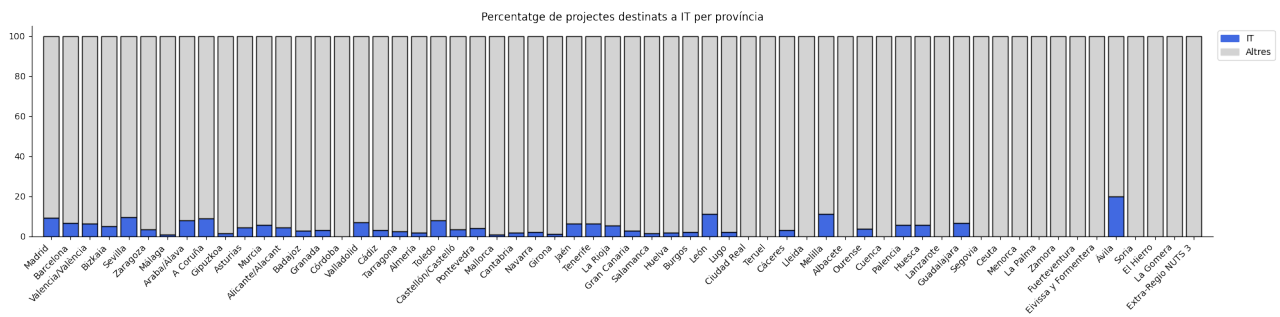
Nom Variable	Descripció	Escala
ID_NOTICE_CN	Identificador únic de la contractació	Caràcter
TED_NOTICE_URL	Url a tota la informació de la contractació	Caràcter
CAE_NAME	Nom oficial del projecte	Caràcter
CAE_TOWN	Ciutat que presenta la contractació pública	Caràcter
CAE_POSTAL_CODE	Url a tota la informació de la contractació	Caràcter
CAE_TYPE	Tipus d'autoritat que contracta (ministeri, entitat regional, etc.	Caràcter
MAIN_ACTIVITY	Activitat principal	Caràcter
TYPE_OF_CONTRACT	Tipus de contracte (S="Serveis", U="Subministraments", W="Feina")	Caràcter
CPV	Codi per especificar l'objectiu principal del contracte	Caràcter
VALUE_EURO_FIN_1	Import en Euros del contracte	Numèric
DURATION	Durada contractació	Numèric
NUTS3	Codi a completar amb la geo-localització	Caràcter

Per tal d'obtenir la geo-localització dels projectes dels quals s'ha obtingut informació completa del camp NUTS3, s'ha usat el diccionari *nuts_dictionary.pickle*, generat a partir de la taula que es troba a la següent pàgina web: <https://ec.europa.eu/eurostat/web/nuts/background>

En aquest gràfic de barres podem observar de manera ordenada d'esquerre a dreta les províncies espanyoles que tenen més projectes amb contractacions públiques europees.



En el següent gràfic de barres podem identificar quin és el percentatge del total de projectes els quals es troben dins del camp IT per província.



En aquesta visualització, podem veure el percentatge de projectes relatius a IT per cada zona NUTS3 a Espanya.



Obtenció de les dades: Web Scraping

Per fer aquest projecte s'ha requerit de Web scraping degut a la gran quantitat d'accessos a documents que requereix, a més de la cerca en aquests de la informació necessària sobre el NUTS3.

El codi usat pel Web scraping consta de dos arguments, un input, el qual és el dataset descrit anteriorment en format de taula (csv) i un nom per l'output que es generarà.

El programa llegeix fila per fila la taula, extraient el URL que es troba a la columna

TED_NOTICE_URL. S'accedeix a aquest a través de request usant un diccionari de capçalera. Aquest diccionari serveix per tal que el codi sembli ser un buscador real i així evitem el bloqueig de l'accés a la pàgina web. Després, usant BeautifulSoup accedirem al contingut de la pàgina i buscarem la classe nutsCode la qual serà guardada a la taula, si es troba disponible. Un cop extreta aquesta informació, de manera aleatòria tindrà lloc un retard per tal de garantir una velocitat continua de les peticions de request, i no saturar la pàgina web. A més, aquests retards també ens garanteixen que no se'ns prohibeixi l'accés a la pàgina web.

5. Agraïments

Com hem descrit anteriorment, les dades obtingudes són propietat de la plataforma Tenders Electronic Daily, la versió online del 'Supplement to the Official Journal of the EU' que ofereix informació sobre les contractacions públiques europees i gestiona a l'any, prop de 746.000 contractacions que mouen al voltant de 345 bilions d'euros.

Aquestes dades han sigut objecte de molts estudis i anàlisis per diferents àmbits, sigui directament sobre aquestes mateixes dades, o com a complement d'informació per altres conjunts de dades.

De fet, la Comissió Europea, expressament divulga l'interès que té en que aquestes dades es reutilitzin a la recerca, tal com diu textualment: "The European Commission is interested in the results of research on public procurement coming from the re-use of this data" i, encara més, manté una wiki amb idees de recerca relacionades amb les seves dades: "To support the exchange of ideas, especially between practitioners and academics, the Commission hosts an open wiki with ideas for research questions and existing examples of reuse related to this dataset. En aquesta wiki podem trobar una quantitat enorme de vincles a interessos de recerca relacionats amb les dades que TED ofereix. En concret, Digital Whistleblower, (DIGIWHIST), recull molts dels estudis que s'han fet en relació a les contractacions públiques.

Respecte el nostre objectiu d'estudi, les contractacions relatives a IT i, per tant, relacionades amb la transformació digital i que utilitzen d'una o altra manera, les dades de TED, hi ha moltes publicacions recents sobre creació de valor en el mercat internacional (Muñoz-Garcia, C., Vila, J., 2019), sobre les contractacions públiques digitals orientades a les dades (Sanchez-Graells, A., 2019) o relacionades amb les polítiques públiques també orientades a les dades (Halsbenning, S., Niemann, M., 2019)

1. European Union comission. (2020). *TED Tenders Electronic Daily*. Recuperat de: <https://ted.europa.eu/TED/main/HomePage.do>
2. Subirats, L. Calvo, M. (2018). Web Scraping. Editorial UOC.
3. Muñoz-Garcia, C., Vila, J. (2019). Value creation in the international public procurement market: In search of springbok firms. *Journal of Business Research*, (101), 516-521. Recuperat de: <https://doi.org/10.1016/j.jbusres.2018.12.041>.

4. Sanchez-Graells, A. (2019). Data-Driven and Digital Procurement Governance: Revisiting Two Well-Known Elephant Tales. *SSRN*, 1-18. Recuperat de: <http://dx.doi.org/10.2139/ssrn.3440552>
5. Halsbenning, S., Niemann, M. (2019). The European Procurement Dilemma-First Steps to Introduce Data-Driven Policy-Making in Public Procurement. *IEEE 21st Conference on Business Informatics (CBI)*. 303-311. Recuperat de: <https://ieeexplore.ieee.org/document/8808005>

6. Inspiració

Avui dia es parla de la transformació digital, la indústria 4.0 i la societat 5.0. Espanya ha posat en marxa l'agenda digital 2025, "España Digital 2025", un pla per actuar sobre 10 eixos principals per encetar la transformació digital del país. Tots aquests conceptes responen a un canvi cultural que està a l'ambient i que pretén ser una millora de present i futur. També, és una necessitat donat el canvi i el dinamisme del món actual. La transformació digital d'una organització permetrà una nova manera d'interrelacionar i transformar el coneixement. A nivell empresarial representa un camí obligatori pel que fa referència a la competitivitat del mercat empresarial global. El terme digital fa referència a la tecnologia que treballa amb dades.

La implementació de tecnologia és un dels aspectes de la transformació digital però no la defineix, en absolut, en si mateixa. Aquesta ha d'estar destinada a la consecució dels objectius de l'organització que pretén fer aquest canvi de paradigma de forma rigorosa i precisa a partir d'uns protocols que guien en aquesta metamorfosi global. Això implica canvis en la forma d'actuar i de pensar en tots els nivells de l'estructura de l'entitat, sent un canvi cultural.

Per tant, és clar que qualsevol estudi relacionat amb la transformació digital pot tenir un interès i una utilitat rellevants.

La motivació d'aquest treball, llavors, neix amb la idea d'explorar, en aquest cas a nivell espanyol, en quin punt estem respecte la transformació digital en relació amb les contractacions públiques de l'àmbit de les tecnologies. I, en concret, com es distribueix aquest concepte a nivell de tot l'estat espanyol. Abastar tot el concepte de Transformació Digital respecte les contractacions de les entitats públiques és molt complex, però, una primera aproximació, pot ser estudiar la distribució dels tipus de projecte (si són relatius a les tecnologies de la informació o no) dins de l'estat espanyol. Aquest és l'objectiu del nostre projecte.

7. Llicència

La llicència que hem escollit per a la publicació d'aquest estudi és la **CC BY-SA 4.0 License**. L'elecció ha estat feta seguint els consells i recomanacions de la comissió europea descrits anteriorment, ja que compleixen les clàusules establertes d'aquesta llicència:

- Es pot compartir, copiar i redistribuir el material lliurement. I a més es pot adaptar i reconstruir el dataset per a qualsevol propòsit, fins i tot el comercial.
- Es té en compte el nom dels creadors del conjunt de dades. D'aquesta manera el treball serà reconegut, podent ser útil proveint informació per a treballs futurs.

- Totes les contribucions futures han de ser publicades sota aquesta llicència. Així es garanteix la distribució del treball seguint els mateixos termes pre-establerts en el treball original.

8. Codi

El codi s'ha desenvolupat en Python i es pot trobar en la branch de codi al repositori de Github.

9. Publicació del dataset

S'ha publicat el dataset en format CSV i amb una breu descripció a Zenodo, obtenint la següent DOI: 10.5281/zenodo.4244682.

10. Contribucions

Contribucions	Signa
Recerca prèvia	Silvia Galan, David Roche
Redacció de les respostes	Silvia Galan, David Roche
Desenvolupament del codi	Silvia Galan, David Roche