

Interfície d'ajuda a la recuperació d'informació a Twitter

Dídac Ritter Grau

Resum—El treball es centre en desenvolupar una interfície gràfica per utilitzar la plataforma Scrapy per obtenir informació de Twitter. Aquesta s'utilitza per extreure, analitzar i realitzar estudis de llenguatge, de tendències polítiques,... a partir de tuits. Per aquesta raó, l'objectiu principal del treball consisteix a millorar la usabilitat d'una interfície d'usuari intuïtiva, "user-friendly", i multiplataforma per tal que qualsevol persona sense coneixements informàtics pugui utilitzar l'API Scrapy aplicada a l'anàlisi de Twitter. Sota aquest requisit apareix la següent pregunta: quins tipus de filtre és el més adient per obtenir informació útil per realitzar l'anàlisi d'informació? La resposta a aquesta pregunta s'obté després de realitzar una entrevista i un anàlisi de requisits amb el client. Per aquesta raó els filtres estan dissenyats i desenvolupats per treballar conjuntament entre ells i així obtenir una informació més precisa. Tenint en compte la investigació i l'anàlisi de requisits, la millor combinació de filtres per a realitzar aquests estudis és un filtre per als usuaris que van dirigits als tuits, conjuntament amb un filtre de proximitat a una ciutat. S'ha obtingut una aplicació funcional que satisfà els requeriments i objectius marcats a l'inici.

Paraules clau— Scrapy per twitter, Crawling, Twitter data, Anàlisi de dades

Abstract— The work focuses on developing a graphical interface to use the Scrapy platform to obtain information from Twitter. This is used to extract, analyze and carry out studies of language, political trends, ... from tweets. For this reason, the main objective of the work is to improve the usability of an intuitive, "user-friendly" user interface, and cross platform so that any person without computer skills can use the Scrapy API applied to the Twitter analysis. Under these requirements a questions appears: What type of filters are needed to obtain useful information? The answer to this question is obtained after an interview and requirements analysis with the client. For this reason the filters are designed to work jointly in order to obtain a better result. Taking into account, the investigation and requirements the most interesting combination to carry out these studies are: A filter a user uses to whom the tweets are addressed to, jointly with the site. A functional application has been obtained that satisfies the requirements and goals set at the beginning

Index Terms— Scrapy for twitter, Crawling, Twitter data, Data analysis.



1 INTRODUCCIÓ

Diàriament es produeix un gran volum d'informació ens les xarxes socials. Aquesta informació és de molta utilitat per als investigadors semàntics i per a la realització d'estudis d'opinió. Per elaborar aquests estudis es necessiten plataformes que siguin capaces de cercar la informació necessària per als investigadors d'una manera senzilla, eficient i eficaç. Malauradament, actualment la majoria d'aquestes aplicacions no estan preparades ni pensades per a que siguin utilitzades per usuaris no experts i, per tant, requereixen de coneixements informàtics. És per aquesta raó que en aquest treball es pretén solucionar aquest problema, creant una interfície d'ajuda per a recuperar dades de Twitter.

En llegir la proposta del treball i fer una avaluació

d'aquest, ràpidament es va veure la importància i el potencial del treball. Cada cop més expresseu opinions a través de les xarxes. En el cas concret del treball ens referim a "Twitter". Aquestes opinions es poden entendre com a informació per poder fer estudis d'opinió i macroanalítica de dades i, d'aquesta manera, extreure conclusions que poden ser molt útils pels diferents tipus d'estudis que realitzen els investigadors. Per altra banda, durant tot el desenvolupament de l'aplicatiu s'ha pogut veure clarament el poder que pot proporcionar la gestió d'aquesta informació, per això sempre s'ha de tenir en compte la finalitat per a la qual s'utilitza aquesta informació.

A continuació, es detallarà l'estudi i el desenvolupament realitzat en totes les seves etapes pel que fa a la gestió de projectes. En segon lloc, s'explicarà la motivació per la qual s'ha realitzat el treball, les necessitats que per les quals s'ha desenvolupat i les oportunitats que ofereix la solució realitzada. Com a segon ítem s'explicarà l'estudi realitzat i la captació de requeriments realitzada per tal d'abordar i desenvolupar una solució òptima i adequada. Seguidament, realitzarem una descripció de la planificació realit-

- E-mail de contacte: didacrittergrau@gmail.com
- Menció realitzada: *Enginyeria del Software*
- Treball tutoritzat per: *Marta Prim Sabrià (Microelectrònica i Sistemes Electrònics)*
- Curs 2018/19

zada, protagonistes i les corresponents fases del treball. Finalment, es realitzaran unes conclusions que aportaran una visió i descripció general del treball.

2 DESCRIPCIÓ

El projecte apareix d'una necessitat del client, la professora Ana Fernández Montraveta del Departament de Filologia Anglesa i de Germanística de la UAB, la qual necessita analitzar i realitzar estudis amb la informació de Twitter. Actualment, això ho realitza amb scripts sovint farragosos i de configuració complexa per això s'urgeix la necessitat d'una interfície gràfica senzilla d'utilitzar.

3 OBJECTIUS I BENEFICIS

Durant aquest apartat es pretén fer una explicació dels objectius marcats durant el treball en funció de la seva rellevància i importància. També s'explicaran els beneficis que aporta el treball al problema plantejat.

3.1 OBJECTIUS PRINCIPALS

Els objectius principals són:

- Facilitar la cerca als usuaris no experts que volen utilitzar la plataforma Scrapy [1].
- Millorar els temps de configuracions i temps de bloqueig per cerques massa grans o poc parametritzades.
- Optimitzar i millorar la gestió que es fa de les dades obtingudes.

3.2 Objectius secundaris

Els objectius secundaris són:

- L'aplicatiu ha de ser multiplataforma. Amb això ens referim a que qualsevol usuari ha de tenir accés indistintament amb independència del dispositiu amb el qual es connecti.
- La informació ha de ser exportable en format *csv* per a que l'usuari pugui treballar amb ella en altres aplicatius.
- L'aplicatiu ha de donar *feed-back* de les accions que està realitzant. En tot moment l'usuari ha de veure activitat per saber que s'està realitzant alguna acció. Aquest objectiu és especialment important, ja que molts processos poden ser llargs i l'usuari podria pensar que l'aplicatiu ha quedat bloquejat.

3.3 Beneficis

Els principals beneficis de desenvolupar aquest projecte són:

- Donar accés a usuaris no experts en software que poden ser d'utilitat pel seu desenvolupament laboral.

- Facilitar la gestió de la informació obtinguda per així millorar el rendiment i el nivell dels estudis.

4 PLANIFICACIÓ I METODOLOGIA

Durant aquest apartat s'exposarà la planificació de les tasques realitzades per desenvolupar el projecte [Taula 1].

Taula 1: Tasques

Tasca	Data inicial	Data finalització
Diagrama d'activitats	1 de març	1 de març
Diagrama de seqüència	1 de març	1 de març
Disseny formulari html	2 de març	3 de març
Disseny base de dades	2 de març	3 de març
Mock-up interfície gràfica	2 de març	3 de març
Instal·lació mòdul MongoDB -PHP	29 de març	31 de març
Programació del disseny formulari html	29 de març	29 de març
Integració de la web amb API d'scrappy	5 d'abril	7 d'abril
Desenvolupament filtres (usuari, temps, paraula i hashtag)	5 d'abril	7 d'abril
Desenvolupament consultes MongoDB	8 d'abril	8 d'abril
Traducció consultes MongoDB a MongoDB - PHP	8 d'abril	8 d'abril
Desenvolupament taula resultat de cerca	9 d'abril	9 d'abril
Refactorització del codi per petició POST AJAX	10 de maig	10 de maig
Desenvolupament <i>.gif loading</i>	11 de maig	11 de maig
Consultar històric	12 de maig	12 de maig
Desenvolupament filtres (proximitat ciutat)	17 de maig	19 de maig
Test inputs	Continuat	Continuat

Tasca	Data inicial	Data finalització
Avaluació noves propostes	15 de maig	16 de maig
Desenvolupament filtres (cerca entre dates)	17 de maig	19 de maig
Desenvolupament export <i>csv</i> .	17 de maig	19 de maig
User testing	25 de maig	26 de maig

Per a gestionar i desenvolupar les tasques de presentades a la Taula 1 s'ha utilitzat la metodologia anomenada "cascada"[5]. Aquesta metodologia conté 6 etapes bàsiques mostrades en la següent imatge (Fig. 1):

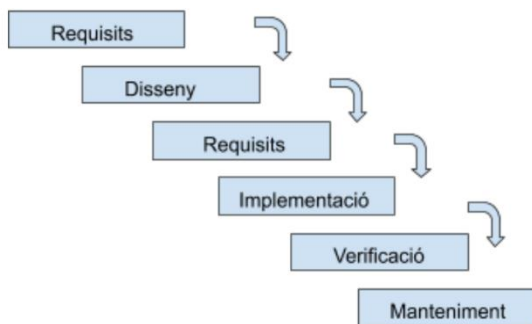


Fig. 1- Metodologia en cascada

La principal premissa d'aquesta metodologia és que no es comença la següent etapa fins que s'ha finalitzat l'etapa anterior.

S'ha escollit aquesta metodologia i no una altra de tipus "Agile"[6], que dins del paradigma actual s'adeqüen més al desenvolupament del software, ja que aquest projecte es desenvolupa individualment i les metodologies "Agile" estan pensades per desenvolupaments en grup.

4.1 EXEMPLE METODOLOGIA

En aquest apartat s'exposarà un exemple de com s'ha utilitzat la metodologia en cascada amb una tasca en concret. La tasca "Desenvolupament dels filtres (usuari, temps, paraula i hashtag)" s'ha desenvolupat de la següent manera:

- Prerequisits: s'han de tenir les tasques anteriors ja finalitzades i testejades.
- Disseny: Estructura lògica (*if-else*) per construir l'*string* (*query*) amb la informació per a la cerca.
- Implementació: Desenvolupament del codi encarregat de construir l'*String* a partir de les dades introduïdes

per l'usuari.

- Proves: Realització de *test inputs* per possibles combinatòries que pot introduir que l'usuari tingui el resultat esperat (introducció del nom d'usuari sense @, introducció de *hashtag* sense el caràcter #, etc.).
- Integració: Es prova el conjunt del codi amb la part del *client-side* per avaluar que el seu comportament és correcte dins del tot el conjunt.

5 ESTAT DEL ART

Realitzant una cerca a internet s'han trobat diferents aplicacions que permeten fer crawling de les webs. Seguidament es descriuran algunes de les més importants i alhora diferents, tant en la seva utilitat com en el nivell de coneixement que es necessita per a la seva correcta utilització.

- BeautifulSoup[2] és una llibreria en Python igual que Scrapy. És una llibreria open-source i gratuïta enfocada a fer crawling de dades dins una web. La instal·lació és igual que la de Scrapy (a partir de l'aplicatiu pip). La seva utilització és semblant a Scrapy, s'han de configurar una sèrie d'Scripts Python per tal de determinar quina és la web de la qual es vol obtenir la informació i que es vol analitzar dins de la web (tags, html, tags css, headers...). Aquesta llibreria també serviria per desenvolupar el treball.

- Semrush[3] és una web que permet fer crawling i analítica de dades d'una web en concret. Aquesta web, a diferència de Scrapy, es pot utilitzar en l'àmbit d'usuari, és a dir, no cal tenir coneixements de programació. Per utilitzar-la només hem d'entrar a la web i escriure quina pàgina es vol analitzar. Semrush ens mostrarà una sèrie d'estadística i backlinks en format dashboard. És una aplicació de pagament però ens permet utilitzar-la en versió de prova durant set dies. Aquest aplicatiu no serviria per solucionar el problema plantejat en aquest treball però entra dins les aplicacions de crawling i analítica de dades web.

- Screaming Frog[4] és un instal·lable que ens permet veure tots els textos que tenen les pàgines web. Això ens permet veure quines són les paraules més rellevants i quines són les més utilitzades, ja que les classifica per nivell d'importància, és a dir, mostra si són header, tags secundaris, descripcions, etc. Aquest aplicatiu no serviria per solucionar el treball però, com Semrush, entra dins les aplicacions de crawling i analítica de dades web i ens permet estudiar la semàntica de les diferents webs.

6 ENTORN DE DESENVOLUPAMENT UTILITZAT

Com s'ha especificat anteriorment, l'aplicatiu ha de ser multiplataforma. Per aquest motiu, s'ha decidit desenvolupar l'aplicatiu en format web, ja que és la manera més senzilla per a què sigui multiplataforma.

La part de *server-side* s'ha desenvolupat amb PHP, per al *client-side* s'ha utilitzat html, AJAX i JavaScript. Per a la part de base de dades s'ha utilitzat una base de dades no relacional MongoDB per no tenir problemes de concurrència a l'hora de fer inserts. MongoDB s'utilitza a partir d'un mòdul que s'instal·la al servidor PHP i d'aquesta manera permet fer les consultes a la base de dades.

7 ARQUITECTURA I DISSENY

7.1 ARQUITECTURA DEL SOFTWARE

Per a realitzar la codificació de l'aplicatiu primerament s'ha realitzat un estudi i disseny de les funcionalitats de com l'usuari interaccionarà amb l'aplicació. Aquestes funcionalitats es veuen reflectides des dels seus diferents punts de vista en els següents diagrames:

El Diagrama 1 correspon al de casos d'ús. Mostra una visió de les accions que ha de dur a terme cada actor dins del software.

El Diagrama 2 correspon al d'activitat. Mostra les accions i comportaments que ha de realitzar l'usuari amb el software per realitzar una cerca.

El Diagrama 3 correspon al de seqüència. Mostra el comportament que realitzarà el software quan un usuari realitza una petició de cerca.

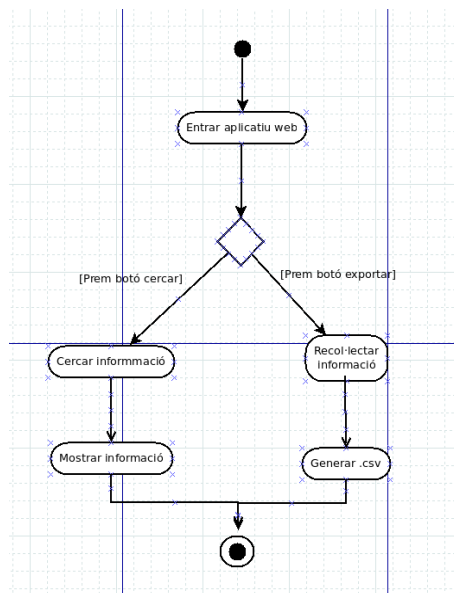


Diagrama.2- Activitats

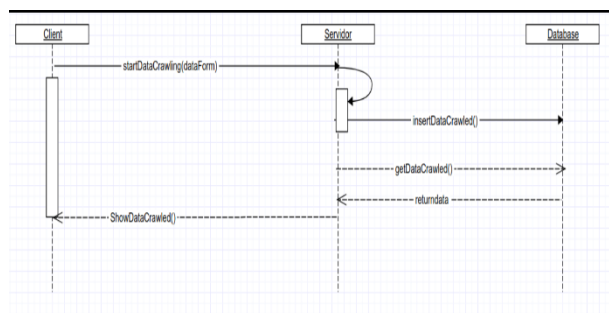


Diagrama.3- Seqüència

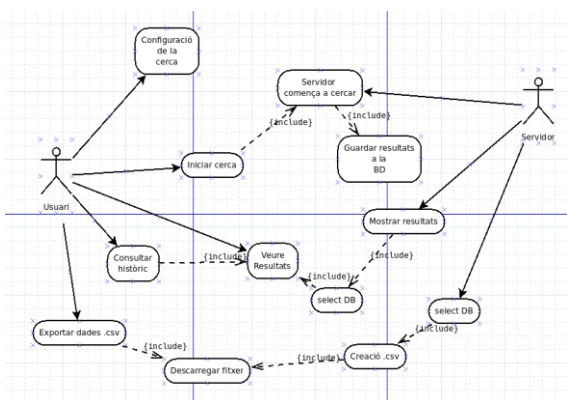


Diagrama.1- Casos d'ús

7.2 ARQUITECTURA DEL CODI

El codi de la web està estructurat basant-se en l'arquitectura MVC (*Moldel view controller*). Aquest tipus d'arquitectura consisteix en separar les accions del *server-side* & *client-side* per tal de que les modificacions posteriors del codi siguin més senzilles de realitzar.

- View: Correspon a la part que interacciona amb l'usuari. Està format pels següents fitxers (*client-side*):

- index.php
- table.php

- Controller: Correspon a la part que interacciona i realitza les accions amb el servidor (*server-side*):

- crawling.php
- historic.php
- exportcsv.php

- External: Correspon a la part del codi en Python encarregada de fer el scraping de la informació mitjançant la plataforma scrapy i arxius necessaris per a què funcioni el mòdul MongoDB [7]:

- Tweepscraper [8]
- Vendo

8 FUNCIONAMENT DE L'APLICACIÓ

L'aplicació funciona d'una forma molt senzilla. L'usuari ha d'omplir els camps del formulari principal. Aquests camps corresponen als filtres que s'utilitzaran per parametritzar la cerca (usuari, temps, dates...). Tots aquests filtres es poden configurar i combinar per obtenir un millor resultat de cerca.

A continuació s'explica el funcionament de l'aplicació a partir d'uns exemples generals d'utilització dels filtres. Aquestes accions corresponen a les més comunes realitzades pels usuaris.

8.1 INTERFÍCIE PRINCIPAL

La Fig.2 mostra el formulari principal de l'aplicació. L'usuari utilitza aquest formulari per parametritzar la cerca dels tuits. Cada camp correspon a un filtre i es combinable amb qualsevol dels altres per tal d'obtenir una cerca més específica.

Fig.2- Formulari de configuració

8.2 CERCA PER LOCALITZACIÓ

La Fig.3 mostra la configuració d'una cerca parametritzada a partir de:

- Nom d'usuari: @DidacRitterGau
- Ciutat: barcelona
- Distància: 20

Aquesta configuració ens retornarà els tuits on aparegui l'usuari @DidacRitterGau que s'hagin realitzat en un radi de 20 km de la ciutat de Barcelona. El resultat d'aquesta cerca es mostra en la Fig.4.

Fig.3- Formulari configuració cerca amb localització

INFORMACIÓ TROBADA		
Nom d'usuari	Text	Hora
DidacRitterGau	@XeniaCaballeroG @laura_aranda3 @Yoel_S94 . Monologos en la playta pic.twitter.com/9zX17vei	2012-07-18 22:34:13

Fig.4- Resultat cerca amb localització

8.3 CERCA USUARI DIRIGIT

La Fig.5 mostra la configuració d'una cerca parametritzada a partir de:

- Tuit dirigit a: 3gerardpique
- Temps: 40
- Cerca des de: 2019-01-01
- Fins a: 2019-03-01

Aquesta configuració retornarà els tuits que han anat dirigits a l'usuari 3gerardpique entre les dates 2019-01-01 i 2019-03-01 (any - mes - dia) i com és un usuari amb molta activitat li afegim un temps de cerca de 40 segons. Part del resultat es mostra a la Fig.6

Fig.5 – Configuració cerca entre dates

usuari	Text del tuit	Data
prakashmurnani	The main knows where and when to strike. Be prepared for some more fireworks... well done @3gerardpique https://twitter.com/CatalansForLes/status/11011917907498824	2019-02-28 23:21:32
alordiezbleton	Sincerament, no he vist un tío mas gilipollas que este, en mi vida! Se puede ser más sopollas?? Al margen del comentario, absolutamente sesgado, a que viene?? Eres tonto, del "o" y "ya" siempre!!	2019-02-28 21:39:30
sergio_privado	And to Nepal.	2019-02-28 21:38:11
Enneraetkellaw	Less little tattle and bar gossiping about what's not happening/never happened in Catalunya and more transparency and diffusion of what's happening in SPAIN is an idea to be applauded! @3gerardpique https://twitter.com/arg_silvaniz/status/110089739590582848	2019-02-28 20:39:07
ignasiros	100% Potser classificada no ha estat el mot més allert, tampoc en en to sarcàstic and divertit, totes les accions son vàlides i importants però no per totom tenen les mateixes conseqüències. El q hem fet i farem ha hem fet entre totthg, però en ha semblat molt bé el teu apunt.	2019-02-28 20:15:08
lascomar	If the players devoted all time to sports and not to do politics or to disrespect the Courts of Justice or to the most citizens of the country... the sport would not get dirty!.	2019-02-28 19:52:24
rival_somozza	Solo to los independentista 7nos retrasados, de como os comen el cocos y vira la republica española esa si existe, hay q tener más huevos y no huir como una rata como parte de ustedes se an ido	2019-02-28 19:51:29
inaki_Trainorral	@zaldigi mira estall! @oraygado es to es un @4000comentarios @uccolapoppe @3gerardpique crack!!	2019-02-18

Fig.6 – Resultat cerca entre dates

8.4 EXPORTAR LES DADES

La Fig.7 i Fig.8 mostren com és el procés d'exportar les dades. Quan es prem el botó exportar apareix una finestra per descarregar el fitxer en format .csv i poder treballar les dades en altres softwares.

Com mostra la Fig.6 el nom del fitxer es posa de forma automàtica la data en el dia que s'ha fet l'export, facilitant així la gestió de tots el fitxers descarregats. La Fig.7 mostra el fitxer .csv exportat. El fitxer conté tres columnes (nom d'usuari, tuit i data)

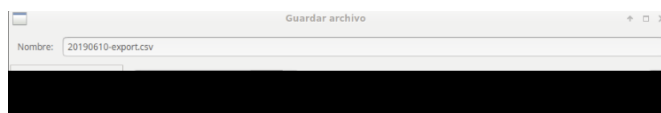


Fig.7 – Finestra guardar fitxer

usuari	tuit	Data
DickRitterGau	@VenerCabalerG @aura_arand3 @lvel_594 Monday en la playa pic.twitter.com/0x17vel	
prakashmurnani	The main knows where and when to strike. Be prepared for some more fireworks... well done @3gerardpique https://twitter.com/CatalansForLes/status/11011917907498824	
alordiezbleton	Sincerament, no he vist un tío mas gilipollas que este, en mi vida! Se puede ser más sopollas?? Al margen del comentario, absolutamente sesgado, a que viene?? Eres tonto, del "o" y "ya" siempre!!	
sergio_privado	And to Nepal.	
Enneraetkellaw	Less little tattle and bar gossiping about what's not happening/never happened in Catalunya and more transparency and diffusion of what's happening in SPAIN is an idea to be applauded! @3gerardpique https://twitter.com/arg_silvaniz/status/110089739590582848	
ignasiros	100% Potser classificada no ha estat el mot més allert, tampoc en en to sarcàstic and divertit, totes les accions son vàlides i importants però no per totom tenen les mateixes conseqüències. El q hem fet i farem ha hem fet entre totthg, però en ha semblat molt bé el teu apunt.	
lascomar	If the players devoted all time to sports and not to do politics or to disrespect the Courts of Justice or to the most citizens of the country... the sport would not get dirty!.	
rival_somozza	Solo to los independentista 7nos retrasados, de como os comen el cocos y vira la republica española esa si existe, hay q tener más huevos y no huir como una rata como parte de ustedes se an ido	
joan_bergues1	@zaldigi mira estall! @oraygado es to es un @4000comentarios @uccolapoppe @3gerardpique crack!!	
McWester12	Would love to! Get ahold to... Happy to take donations! Always been on my bucket list to see it and the You Camp!	
carloscuar	What for? In cataland there are over 50k neomercos, but safe declared by social separatist, armed demonstrations by organized neomercos every day, constant threats to other 50k people by neomercos	

Fig.8 – Fitxer exportat

8.5 CONSULTAR HISTÒRIC

La Fig.9 mostra el resultat de prémer el botó d'històric. Aquesta acció realitza una consulta instantània a la base de dades i retorna tot el seu contingut en format taula web.

namaria_	You know what you have to do in Copa. 2-2 will be enough :)	2019-02-17 21:54:12
adilhmdaul	Back to twitter geri	2019-02-17 21:53:32
gerpep	Pique is correct. Barça seems to play for possession and then other teams press them they lose@ball Nd then have to rush to defend. Sometimes defense short. Barça then ends up behind and fights for a tie.	2019-02-17 20:02:41
isortagi44	Gerard Pique Collectible Figurine to Buy: https://amzn.to/2BFzrZU	2019-02-17 20:01:29
Isortagi44	Gerard Pique Collectible Figurine to Buy: https://amzn.to/2BFzrZU	2019-02-17 20:01:04
Varnosgir11	Please stick to stay away from !!	2019-02-17 19:27:47
DonnyKerabatos2	They play to slow under Valverde. No rhythm. No plays beyond bounce off and back to Lio	2019-02-17 18:51:56
beINSPO RTSUSA	"If we don't speed up the ball, we have possession...and we turn into a weak team." @3gerardpique tells @FCBarcelona what needs to happen to avoid an upset against @OL. http://bit.ly/2EgYXP	2019-02-17 18:35:00
PrimeDembele_	I hope so but we are getting to the latter stages of every competition should have been rested yesterday	2019-02-17 18:01:32

Fig.9 – Taula històric online

9 PROBLEMES SORGITS

9.1 TAULA DINÀMICA

Duran el disseny del software es va pensar en implementar una d'una taula dinàmica per anar mostrant els resultats de forma progressiva. Aquest objectiu estava pensat per solucionar la part del *feed-back* amb l'usuari. No s'ha pogut realitzar a causa de problemes amb la tecnologia de base de dades utilitzada. Altres tecnologies tenen aquesta funcionalitat desenvolupada a partir d'una funció que va retornant valors i guarda en un apuntador l'última tupla retornada. Malauradament, en MongoDB treballant conjuntament amb PHP no està desenvolupada, per tant s'ha de desenvolupar. Per motius del temps en el que esta acotat el treball s'ha donat prioritats a tasques que es consideraven més importants, s'ha canviat la forma de donar el *feed-back* a l'usuari. Finalment s'ha posat un *.gif* perquè l'usuari no perdi la sensació de que s'està realitzant alguna acció i d'aquesta manera seguir donant *feed-back*.

9.2 INSTAL·LACIÓ MÒDUL MONGODB PHP

Durant el procés d'instal·lació del mòdul de mongoDB per a php s'han tingut diversos problemes. La majoria de problemes sorgits han sigut degut a la documentació pobra o desactualitzada. La instal·lació per a Linux ha sigut realment costosa a nivell de temps. Finalment, s'ha resolt sense problemes però s'ha vist que mongoDB no és molt utilitzat per treballar amb php, sinó que s'utilitza més freqüentment a web nodeJS o JavaScript.

10 PROVES DE TEST

Durant tot el projecte s'han realitzat principalment proves de caixa negra. Les proves s'han centrat en fer test dels inputs del formulari web:

- Caixa negra: s'han realitzat proves dels inputs. Els valors han sigut del tipus numèric i *string*.
- Exploratori testing: s'han realitzat proves com prémer botons sense omplir dades i buscant comportament estranys.
- *Test case*: s'han dissenyat diferents *test cases* per a què user-testers els segueixen i puguin donar *feed-back* dels errors trobats. En l'annex A1 és mostra un exemple d'un *test case*.

11 CONCLUSIONS

Els objectius que s'han marcat des del principi del projecte s'han assolit pràcticament en tota la seva totalitat. Com s'ha explicat en el punt 9 han aparegut diferents problemes que han obligat a realitzar canvis. Aquests canvis sempre s'han realitzat de manera estudiada i mai generant renúncia als objectius principals, per això es pot dir que s'ha assolit la fita marcada del projecte.

Durant tot el projecte s'han après molts conceptes interessants tant a nivell personal com a nivell professional com són:

- Desenvolupament de filtres útils per a realitzar estudis d'opinió. Per desenvolupar aquests filtres s'han realitzat estudis respecte a quin és la informació útil que es pot extreure de Twitter i una presa de requeriments amb la professora Ana Fernández Montraveta del departament de Filologia Anglesa i de Germàniques de la UAB que ha proposat el projecte.
- Disseny i desenvolupament d'una interfície gràfica funcional seguint els patrons apresos en diferents assignatures del grau d'enginyeria informàtica.
- Planificació i gestió de les tasques del treball des de zero seguint una metodologia. Sempre que han aparegut noves tasques, s'han solucionat. Sempre que s'ha hagut de prioritzar tasques, s'ha fet de forma correcta. Aquesta prioritització s'ha realitzat per intentar obtenir sempre més funcionalitats i realitzar una aplicació sòlida i alhora útil.

Finalment destacar que s'han après altres conceptes no tant a nivell de tecnologia i informàtica però igual d'importants avui dia. Aquests conceptes corresponen a l'avaluació ètica sobre quin tipus d'estudis es vol realitzar amb les dades obtingudes. A més a més, avui aquest tipus d'estudis corresponen a un tema d'actualitat. Per tant, cal destacar la importància de poder tenir aquesta informació i estudiar-la, però sempre tenint presents valors ètics i quin impacte poden tenir en el món actual.

12 LÍNIES FUTURES

Les principals línies d'evolució d'aquest projecte són:

- Implementació de més filtres per millorar l'obtenció d'informació i que sigui encara més precisa.
- Implementació d'un algoritme d'interpretació de llenguatge per tal de poder avaluar el tarannà dels tuits[9].

AGRAÏMENTS

En primer lloc m'agradaria agrair a Marta Prim Sabrià, la meua tutora del TFG que sempre m'ha atès quan ha sigut necessari amb molt bon tracte i ajudant-me. Per altra banda també agraeixo la reunió que he tingut amb la professora Ana Fernández Montraveta que ha expressat l'interès i satisfacció al veure funcionant l'aplicació i m'ha motivat i explicat possibles línies futures a seguir en el món de tractament de dades i algorítmica semàntica.

BIBLIOGRAFIA

- [1] Scrapy, <https://docs.scrapy.org> [En línia]. Disponible a: <https://docs.scrapy.org/en/latest/>
Consulta: 18/02/2019
- [2] BeautifulSoup, <https://pypi.org/> [En línia]. Disponible a: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>
Consulta: 26/02/2019
- [3] SEMrush, <https://www.semrush.com> [En línia]. Disponible a: https://www.semrush.com/lp/sem/es/?https://www.semrush.com/lp/sem/es/%3Fkw=Semrush&cmp=ES_SRCH_Brand_Semrush_EN&la-bel=Brand_Semrush&network=g&gclid=CjwKCAiA_P3jBRAqEiwA-ZyWWaKK6yN7-5pv_AffIP2Hq66eQ2-7CBYfa-qKVgekEgVkcvmam68MFteBoCrYUQAyD_BwE
Consulta: 26/02/2019
- [4], Scream Frog, <https://www.screamingfrog.co.uk> [En línia]. Disponible a: <https://www.screamingfrog.co.uk/seo-spider/user-guide/>
Consulta: 26/02/2019
- [5], Wikipedia, <https://es.wikipedia.org> [En línia]. Disponible a: https://es.wikipedia.org/wiki/Desarrollo_en_cascada
Consulta: 20/02/2019
- [6], Vanessa Rosselló Villán, Las metodologías ágiles más utilizadas y sus ventajas de la empresa, 3 oct 2018, [En línia]. Disponible a: <https://www.iebschool.com/blog/que-son-metodologias-agiles-agile-scrum/>
Consulta: 20/02/2019
- [7] Manual PHP, <https://www.php.net/manual/es> [En línea]. Disponible a: <https://www.php.net/manual/es/class.mongodb.php>
Consulta: 23/04/2019
- [8] jonbakerfish, <https://github.com/> [En línea]. Disponible a: <https://github.com/jonbakerfish/TweetScraper>
Consulta: 08/04/2019
- [9] Sociedad Española para el procesamiento del lenguaje natural [En línea]. Disponible a: <http://www.sepln.org/news/xxxv-congreso-internacional-de-la-sepln/>
Consulta: 21/05/2019

APÈNDIX

A1. EXEMPLE TEST CASE: EXPORTAR DADES

Test Case ID	TC_ExportData_03		
Description	Exportar les dades de la BD en format <u>.csv</u>		
Module	Export data from server		
Prepared By	Didac Ritter	Date Prepared	3 de juny 2019
Reviewed / Updated	Didac Ritter	Date Reviewed	
Tested By	Xènia Caballero	Date Tested	5 de juny 2019
Test Activities			
Sl. No	Step Description	Expected Results	Actual Results
1	prèmer el botó "Exportar"	Apareix la finestra de gestió de descarregar del navegador	GOOD
2	Mirar que el nom del fitxer que es descarregarà es igual al de la data actual		GOOD
3	prèmer el botó descarregar	El fitxer es descarrega al ordinador	GOOD
4	Obrir el fitxer <u>.csv</u>	Veure que el fitxer no esta corrupte i conte tres columnes. (Nom d'usuari, text, data)	GOOD