

Creació d'una app de Chatbot integrable a App de missatgeria i apps webs

Carlos Valero Pérez

Resum—D'onada la gran quantitat de demanda d'informació que els centres educatius han d'assumir cada cop que comença un nou curs acadèmic, sorgeix la necessitat de facilitar la feina a les persones encarregades d'atendre aquesta demanda. Els chatbots, també coneguts com assistents virtuals, són capaços de simular mantenir una conversa amb un ésser humà donant respostes automatitzades als dubtes plantejats. Aquest treball de fi de grau té com objectiu la realització d'una prova de concepte per generar un prototip d'assistent virtual que pugi atendre la demanda d'informació sobre els estudis oferts per la UAB, i proporcionar de manera automatitzada la informació en funció de les dades indicades pels usuaris. Per tal d'assolir aquest objectiu, en primer lloc s'ha desenvolupat una aplicació per obtenir les dades del web de la UAB fent servir tècniques de scraping i mineria de dades. Després, per tal de donar persistència a les dades tractades, s'ha fet servir una base de dades MongoDB de tipus NoSQL. Finalment s'ha desenvolupat un chatbot amb Microsoft Bot Framework SDK per a Node.js i un algoritme d'aprenentatge profund que permet adaptar els resultats en funció de les dades proporcionades pels usuaris.

Paraules clau—Prova de concepte, assistent virtual, chatbot, web scraper, mineria de dades, MongoDB, NoSQL, Bot Framework, Node.js, aprenentatge profund, aprenentatge computacional.

Abstract— Given the large amount of demand for information that schools must assume each time a new academic year begins, the need to facilitate the work of those responsible for meeting this demand arises. Chatbots, also known as virtual assistants, are able to simulate a conversation with a human being by giving automated answers to the questions raised. The aim of this end-of-degree project is to carry out a proof of concept to generate a prototype of a virtual assistant to meet the demand for information on the studies offered by the UAB, and to provide the information in an automated way according to the data indicated by the users. To achieve this objective, firstly an application has been developed to obtain the data from the UAB website using scraping and data mining techniques. Then, in order to give persistence to the treated data, a MongoDB database of type NoSQL has been used. Finally, a chatbot has been developed with Microsoft Bot Framework SDK for Node.js and a deep learning algorithm that allows to adapt the results according to the data provided by the users.

Index Terms—Proof of concept, virtual assistant, chatbot, web scraper, data mining, MongoDB, NoSQL, Bot Framework, Node.js, deep learning, machine learning.



1 INTRODUCCIÓ

ELS chatbots, també coneguts com assistents virtuals, són cada cop més freqüents dins el món empresarial.

Basats en intel·ligència artificial, són capaços de simular mantenir una conversa amb un ésser humà donant respostes automatitzades als dubtes plantejats. Aquest treball de fi de grau consisteix en la realització d'una prova de concepte per generar un prototip d'assistent virtual configurable fàcilment i integrable a diferents solucions tecnològiques. L'objectiu principal és desenvolupar un chatbot que pugi atendre la demanda d'informació sobre els estudis oferts per un centre educatiu, en aquest cas la UAB, i proporcionar de manera automatitzada la informació en funció de les dades indicades pels usuaris. A més es guarda la informació sobre les

consultes dels usuaris per a ser tinguda en compte per posteriors consultes. En primer lloc s'ha realitzat un estudi sobre els diferents entorns de desenvolupament de chatbots disponibles i s'ha triat el més adient tenint en compte les necessitats a tractar. Un cop escollit l'entorn de treball s'ha fet un anàlisi de les dades a tractar, com obtenir-les i com emmagatzemar-les pel seu posterior ús. Finalment s'ha realitzat el desenvolupament de l'aplicació. Aquest document està estructurat en les seccions definides a continuació. En primer lloc es presenta el problema a tractar, seguidament es plantegen els objectius a assolir durant la realització del projecte, la situació actual o estat de l'art en que es troben aplicacions orientades a resoldre problemes similars en l'àmbit de la educació, la metodologia utilitzada en conjunt amb la planificació i desenvolupament del treball realitzat, els resultats obtinguts un cop realitzat el desenvolupament, finalment s'exposen les conclusions extretes un cop finalitzat el projecte i la bibliografia de referència.

- E-mail de contacte: carlos.valero@e-campus.uab.cat
- Menció realitzada: Tecnologies de la Informació.
- Treball tutoritzat per: Joan Giner (dEIC)
- Curs 2019/20

2 OBJECTIUS

L'objectiu d'aquest projecte és desenvolupar un assistent virtual que sigui capaç de proporcionar informació sobre els diferents cursos oferts per la UAB. A partir d'un conjunt de dades obtingudes del web i de les dades que l'assistent sol·licita a l'usuari, ja sigui amb preguntes explícites o amb tries que l'usuari ha de fer sobre les opcions proposades per l'assistent, aquest retornarà informació més detallada sobre l'opció final seleccionada, a més de guardar la informació aportada per l'usuari per a tenir-la en compte en posteriors cerques. D'aquesta manera, les opcions més seleccionades tindran prioritat a l'hora de ser proposades en futures consultes d'altres usuaris.

Per tal d'assolir aquest objectiu general sorgeixen altres objectius específics a tractar. En primer lloc cal definir l'entorn de treball a utilitzar en el desenvolupament de l'assistent, tenint en compte les diferents alternatives disponibles.

Cal definir també la font d'informació des d'on s'extrauran les dades que l'assistent tractarà i després comunicarà a l'usuari, i com s'emmagatzemen aquestes dades pel seu posterior us.

Seguidament s'ha de definir el conjunt de preguntes que l'assistent realitzarà a l'usuari així com generar opcions de sol·licitud d'elecció que l'assistent oferirà a l'usuari.

Finalment definir el mètode per a tenir en compte les tries finals que es van realitzant per a ser tingudes en compte en usuaris posteriors.

3 ESTAT DE L'ART

Els chatbots són aplicacions software capaços de simular mantenir una conversa amb un ésser humà, tant per escrit com de forma oral, donant respostes automatitzades als dubtes plantejats. També coneguts com assistents virtuals, [1] els chatbots tenen com a característica principal l'ús de tècniques basades en la intel·ligència artificial, que fa referència a la seva capacitat per a interpretar les respostes i decidir segons aquestes. El terme "ChatterBot" va ser emprat per primera vegada al 1994 per Michael Mauldin al seu article "ChatterBots, TinyMuds, and the Turing Test: Entering the Loebner Prize Competition" per descriure aquest tipus de programes de conversa. [2] Tot i això, es considera que el primer chatbot com a tal es va crear al 1964 al MIT, anomenat Eliza, busca paraules clau en el text introduït per l'usuari i respon amb un model registrat a la seva base de dades. Mes de trenta anys després, al 1996, es va desenvolupar el chatbot Alice, basat en processament de llenguatge natural, fa servir tècniques de reconeixement de patrons per tal de trobar la resposta més apropiada a l'entrada de l'usuari. Al 2008 es va presentar Cleverbot que, a diferència d'altres chatbots, les seves respostes no estan preprogramades sinó que apren directament de la informació humana, on l'usuari escriu un comentari o pregunta i Cleverbot troba totes les paraules clau o una frase exacta que coincideixi amb l'entrada. [3] Actualment els chatbots es fan servir en gran mesura en el sector de l'atenció al client però també es poden

trobar en altres àmbits com l'entreteniment, el sector bancari, l'atenció sanitària i l'educació. Google Assistant, Alexa de Amazon i Siri de Apple son també exemples d'assistents virtuals àmpliament coneguts. [4] [5]

Segons l'article publicat per Rainer Winkler i Matthias Söllner al 2018, el notori augment dels chatbots en els darrers anys permet fer una classificació d'aquets segons tres dimensions diferenciades. La primera "plantejaments de construcció" distingeix entre models basats en recuperació i models generatius. Els models basats en recuperació utilitzen un repositori de respostes predefinides i un algoritme per escollir una resposta apropiada segons la entrada i el context. Aquest algorisme pot ser molt simple, com la coincidència d'expressions basades en regles, o més complex, com els classificadors basats en aprenentatge computacional. Els models generatius no responen amb respostes predefinides sinó que generen respostes a partir de la entrada amb l'ajuda de tècniques d'aprenentatge computacional. Ambdós enfoc tenen avantatges i inconvenients. Per una banda, els mètodes basats en recuperació són més fàcils de construir i no cometen errors gramaticals, d'altra banda, no poden tractar casos a on no existeix una resposta predefinida. Els models generatius son capaços de remetre a la informació de entitats contextuais mencionades anteriorment durant la conversa. Això permet que el chatbot tingui una aparença més humana i sigui capaç de mantenir diàlegs més extensos.

A diferència dels sites web i les aplicacions, els chatbots poden anticipar-se a les consultes dels usuaris i, per tant, interactuar proactivament amb ells mostrant exactament la informació que estan buscant, de manera que els usuaris no hagin d'iniciar explícitament una conversa amb assistents humans o buscar la informació requerida dins del contingut complet del site web. L'ús de chatbots permeten augmentar la satisfacció de l'usuari reduint els temps de resposta i oferint disponibilitat 24 hores al dia, a més, permeten estalviar costos als serveis d'atenció al client reduint el personal humà dedicat. Els chatbots són també una gran eina d'anàlisi de negoci, ja que les converses entre usuaris i chatbots es poden analitzar automàticament per determinar millor els requisits dels clients i, per tant, oferir millors productes i serveis. Encara que els chatbots poden representar una molt bona alternativa de cara a l'atenció a l'usuari, per molta inversió que es realitzi per millorar els assistents virtuals, sempre existeix la possibilitat de error o la possibilitat de presentar qüestions que una màquina no sigui capaç de resoldre. En aquest sentit, la resposta davant d'imprevistos continua sent un tema de humans, es per això que els chatbots s'han de fer servir com a complement i no pas com una substitució. [6]

Pel que fa a l'àmbit de la educació, s'han desenvolupat ja diversos chatbots que permeten abordar diferents temes dins del sector. Per exemple, la universitat de Georgia ha creat un chatbot basat en la plataforma Watson de IBM anomenat "Jill Watson" desenvolupat específicament per tractar els missatges del fòrum dels estudiants matriculats

a un curs d'informàtica. En especial al àmbits de aprenentatge a gran escala en les universitats o en cursos massius online tenen gran potencial a l'hora de compensar la falta de personal docent per tal d'aportar un tracte personalitzat als alumnes. [3]

Un altre cas interessant és "Lola", un chatbot desenvolupat a la Universitat de Murcia basat en la plataforma DialogFlow de Google. Aquest chatbot es capaç d'atendre a joves estudiants que, un cop acabades les proves d'accés a la universitat, es dirigeixen massivament i en pocs dies als punts d'informació de les universitats a on pretenen matricular-se. [7] [8]

4 METODOLOGIA

Per a la realització d'aquest treball de final de grau s'han establert un conjunt de sessions de control per part del tutor del projecte per revisar el progrés del desenvolupament i el contingut dels informes dels lliuraments pertinents.

Lliurament	Data	Descripció
Informe Inicial	08/03/2020	· Definició del projecte · Objectius · Planificació
Informe de Progrés I	19/04/2020	· Primera fase del desenvolupament · Documentació
Informe de Progrés II	24/05/2020	· Revisió de planificació i objectius · Segona fase del desenvolupament · Documentació
Informe Final	14/06/2020	· Revisió de planificació i objectius · Finalització del desenvolupament · Documentació
Dossier i Presentació	28/06/2020	· Redacció del dossier · Elaboració de la presentació

Durant la primera fase de desenvolupament s'ha realitzat la tria de l'entorn de desenvolupament a utilitzar, per això s'ha realitzat un estudi sobre les diferents alternatives disponibles. S'ha definit també la font d'informació des d'on s'extrauran les dades que el chatbot tractarà i comunicarà a l'usuari. Finalment, un cop s'han obtingut les dades necessàries, el següent punt a tractar consisteix en fer arribar aquestes dades a l'assistent virtual.

Durant la segona fase del desenvolupament es defineix la manera en la que s'emmagatzema la informació obtinguda del web i com l'assistent accedirà a aquesta informació. Durant aquesta segona fase de desenvolupament és on s'ha realitzat també el desenvolupament del chatbot.

Durant la darrera fase del desenvolupament s'ha definit el procés pel qual les dades proporcionades per l'usuari afectaran al resultat de futures consultes d'altres usuaris.

5 ESTUDI I DEFINICIÓ DEL FRAMEWORK

Dins el món de la intel·ligència artificial hi ha moltíssimes opcions a triar a l'hora de desenvolupar un assistent virtual. IBM Watson, Google Dialogflow i Microsoft Bot Framework són alguns exemples d'eines disponibles per al desenvolupament. Per tal d'acotar el problema, es fa una tria entre les dues darreres opcions.

· Google Dialogflow

Es tracta d'un framework desenvolupat per Google, abans conegut com api.ai, que permet la creació de chatbots capaços d'entendre el llenguatge natural. [9] La configuració està molt orientada en l'ús de les interfícies gràfiques proporcionades, tot i que molt complertes, deixen més de banda l'ús de codi customitzat per part dels desenvolupadors que el fan servir.

· Microsoft Bot Framework

Es tracta del framework desenvolupat per Microsoft per a la creació de chatbots, a diferència del cas anterior, està més orientat cap a desenvolupadors que busquin una personalització més profunda i unes capacitats més robustes. [10] És precisament aquest el motiu pel qual s'ha escollit aquest darrer framework per al desenvolupament del projecte.

6 FONT D'INFORMACIÓ

El primer punt a tractar durant el desenvolupament consisteix en definir la font d'informació des d'on s'extrauran les dades que el chatbot tractarà i comunicarà a l'usuari. La intenció és que el chatbot sigui capaç de donar informació sobre qualsevol dels estudis que ofereix la UAB. Donat que la UAB no disposa d'una api pública que subministri aquesta informació, ni tampoc disposem d'accés a la base de dades interna a on s'emmagatzema, es plantegen diferents alternatives per tal d'arribar a les dades necessàries. Una opció seria obtenir les dades mitjançant el DDD, Dipòsit Digital de Documents de la UAB, que és l'eina a partir de la qual es recopila, gestiona, difon i preserva la producció científica, docent i institucional de la universitat alhora que recull documents digitals que formen part de les col·leccions de les biblioteques de la UAB. [11] Una altra font d'informació és la pàgina web oficial de la UAB, a on es pot obtenir gran quantitat d'informació sobre els diferents cursos oferts, així com els mètodes d'accés per a poder cursar-los. Tota la informació mostrada a la web està molt ben organitzada tant pel que fa al disseny gràfic com al seu codi font, el que permet simplificar en gran mesura la utilització de tècniques de scraping, que s'engloba dins l'àmbit del data mining o mineria de dades. L'objectiu general del procés de mineria de dades consisteix en extreure informació d'un conjunt de dades i transformar-la en una estructura comprensible per al seu ús posterior.

El scraping és una tècnica utilitzada per a la extracció de dades dels sites web de forma automàtica mitjançant el protocol http, principalment a llocs a on no s'hagi definit una api oficial per a l'obtenció de les dades publicades. Tot i que el procés de scraping es pot realitzar manual-

ment per un usuari, el terme es refereix típicament a processos automatitzats implementats fent servir bots també anomenats web crawlers. El procés en si es tracta d'una forma de copia a on s'identifiquen les dades específiques i es copien del web, normalment a una base de dades o full de càlcul locals, per la seva posterior recuperació o anàlisi. [12]

En aquest cas s'ha desenvolupat una aplicació de tipus web scraper basada en Node.js, que es tracta d'un entorn en temps d'execució, multiplataforma i de codi obert, que permet la execució de codi JavaScript sense la necessitat de fer servir un navegador web. [13] Per tal de obtenir el codi font de la pàgina i poder identificar i extreure la informació necessària s'han fet servir les llibreries indicades mes endavant, que es poden obtenir mitjançant npm, el gestor de paquets de node. Per una banda, s'ha fet servir la llibreria Request-Promise per tal d'enviar peticions a la url de la web i obtenir com a resposta el codi font HTML. D'altra banda, per poder tractar el codi HTML rebut s'ha fet servir la llibreria Cheerio, que es una de les llibreries mes utilitzades ja que permet cercar elements en el codi mitjançant selectors i mètodes de jQuery. [14] [15] A continuació es mostra un exemple de les dades obtingudes pel cas concret del grau en enginyeria informàtica.

title:	"Grau en Enginyeria Informàtica"
description:	"Actualment, els professionals en els reptes socials"
content:	"El grau en Enginyeria Innovacions tecnològiques."
faculty:	"Escola d'Enginyeria"
campus:	"Campus de Bellaterra"
calendar:	"Durada: 4 cursos - 240 crèdits"
places:	"Places: 190"
marks:	"Nota de tall Batxillerat / CFGS: 7,990"
price:	"Preu per crèdit: 39,53 euros."
language:	"Idioma: Català (70 %), optatives, anglès (5 %)."
mode:	"Modalitat: Presencial."

Fig. 1 Informació obtinguda

Finalment, un cop s'han obtingut les dades necessàries, el següent punt a tractar durant el desenvolupament consisteix en fer arribar aquestes dades a l'assistent virtual. Per aconseguir això s'ha desenvolupat una api simplificada fent servir la llibreria Express. Aquesta llibreria proporciona un framework simple i flexible per a aplicacions web dins de Node.js. Disposa de gran varietat de mètodes per treballar amb el protocol http el que permet crear una API de manera ràpida i simple. [16] Tot i que, tal com s'explica més endavant, s'ha descartat tractar la informació al vol a través de peticions get cap a la api i s'ha optat per fer servir un base de dades que emmagatzemi aquesta informació, de manera que el seu tractament sigui més fàcil de realitzar.

A continuació es defineix la manera en la que s'emmagatzema la informació obtinguda del web i com l'assistent accedirà a aquesta informació. Per tal de facilitar l'accés i la consulta de les dades obtingudes del web i, a més, dotar de persistència a aquestes dades, s'ha fet servir MongoDB, que és un sistema de base de dades de tipus NoSQL orientat a documents. En aquest tipus de

base de dades, enlloc de fer servir taules per emmagatzemar la informació, tal i com es fa amb les bases de dades relacionals, la informació es guarda en un format similar a JSON de manera que la integració amb determinades aplicacions resulta més fàcil i ràpida. [17] Per poder emmagatzemar les dades obtingudes del web s'ha modificat el codi de la aplicació de scraper desenvolupada anteriorment per tal d'afegir la part corresponent a l'accés a la base de dades, fent servir funcions asíncrones de Node.js. Les funcions asíncrones són instàncies del constructor AsyncFunction y la paraula reservada await està permesa dins d'aquestes funcions. Aquestes paraules reservades async i await permeten definir un comportament asíncron basat en promises amb un estil més clar, sense la necessitat de configurar explícitament promise chains [18] [19]

A continuació es mostra un diagrama del disseny de la base de dades utilitzada per emmagatzemar i tractar la informació obtinguda a través del procés de scraping del web.

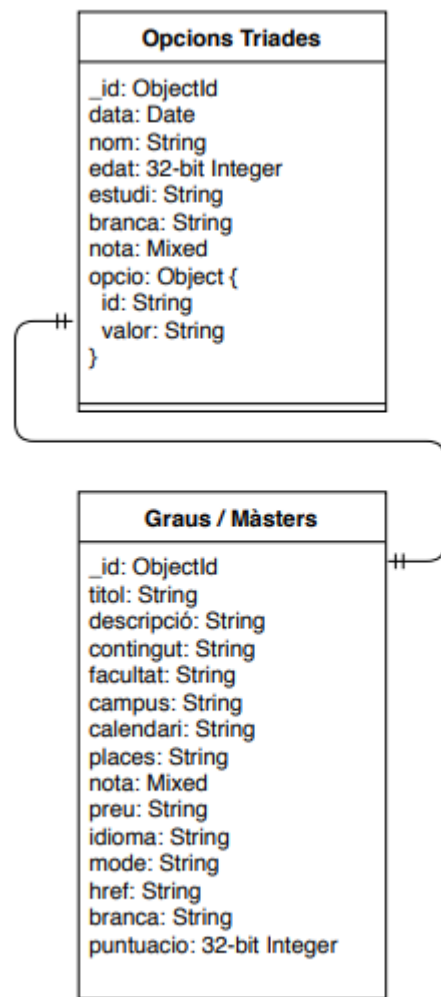


Fig. 2 Diagrama Base de Dades

7 DESENVOLUPAMENT DEL CHATBOT

Per crear el chatbot basat en Microsoft Bot Framework SDK per a Node.js s'ha fet servir el generador Bot Builder de Yeoman, que permet generar automàticament l'estructura del bot mitjançant plantilles predefinides. [20] El generador Yeoman crea una aplicació web de tipus restify a on es construeix el bot. Un bot és una aplicació amb la que els usuaris poden interactuar de forma conversacional, ja sigui mitjançant text escrit, gràfics o veu. Cada interacció entre l'usuari i el bot genera una activitat. Bot Framework Service, que és un component de Azure Bot Service, envia informació entre l'aplicació d'usuari connectada al bot i el propi bot. En una conversa els participants parlen normalment un després de l'altre, per torns. Per Bot Framework SDK, un torn consisteix en la activitat de l'entrada de l'usuari en el bot i qualsevol activitat que el bot retorna a l'usuari com a resposta immediata. Es pot entendre el torn com el processament associat amb l'arribada d'una activitat determinada. L'objecte turn context proporciona informació sobre l'activitat, com el remitent i receptor, el canal i altres dades necessàries per al processament de l'activitat. També permet afegir informació durant el torn en les diferents capes del bot. Turn context és una de les abstraccions més importants al SDK, no només porta l'activitat d'entrada a tots els components de middleware i la lògica de l'aplicació, sinó que també proporciona el mecanisme mitjançant el qual poden enviar activitats de sortida. [21]

Els diàlegs són un concepte central del SDK i proporcionen una forma útil d'administrar converses amb els usuaris. Els diàlegs són estructures d'un bot que actuen com funcions en el programa del bot. Cada diàleg està dissenyat per a realitzar una tasca concreta. Es pot especificar l'ordre de diàlegs individuals per guiar la conversa i es poden cridar de diferents maneres, pot ser en resposta a un usuari o en resposta a altres diàlegs. Els diàlegs són un mètode de implementació d'una conversa amb diferents torns i, com a tal, es basen en un estat persistent durant els diferents torns. [22] En aquest cas s'ha definit un diàleg en cascada per obtenir la informació necessària de l'usuari i poder donar una resposta adient. Un cop mostrada la resposta a l'usuari s'emmagatzema a la base de dades tota la informació proporcionada.

A la columna dreta es mostra un diagrama amb el diàleg que s'ha definit per al bot. En primer lloc es demana a l'usuari que introdueixi el seu nom i, de manera opcional, que introdueixi la seva edat. A continuació es demana a l'usuari que indiqui quins estudis vol realitzar entre les dues opcions disponibles, Grau o Màster. En cas que s'hagi seleccionat la opció Grau es demanarà a l'usuari que introdueixi també la seva nota d'accés. Seguidament es demana a l'usuari que seleccioni una de les diferents branques de coneixement en la que es classifiquen els estudis segons la informació obtinguda des del web. Un cop seleccionada aquesta branca de coneixement es proporciona a l'usuari un llistat de estudis que estan inclosos en la branca seleccionada, en el cas dels graus només es mostren els estudis a on la nota de tall sigui igual o inferior

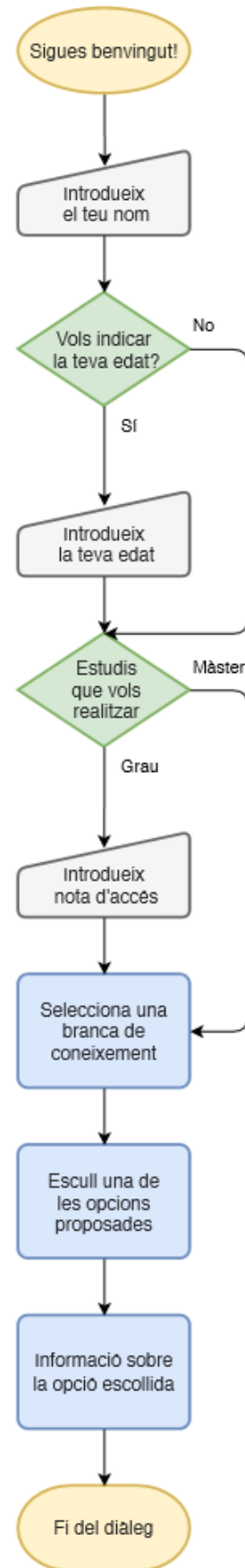


Fig. 3 Diàleg del chatbot

a la nota d'accés introduïda anteriorment per l'usuari i, a més, aquest llistat s'ordena segons la nota de tall de cada curs de major a menor nota. Finalment l'usuari ha de seleccionar un dels cursos mostrats en aquest llistat i l'assistent retornarà informació ampliada sobre aquesta darrera opció seleccionada, a més de un botó amb un enllaç al web on s'inclou tota la informació referent al l'estudi seleccionat. L'assistent guarda també tota la informació introduïda per l'usuari així com les opcions que ha anat seleccionant entre les proposades.

8 ALGORITME DE DEEP LEARNING

Un cop definit el diàleg que es durà a terme, durant la darrera fase del desenvolupament s'ha definit el procés pel qual la selecció final que realitza l'usuari és tinguda en compte per modificar l'ordre en que apareixen les opcions proposades en futures consultes per altres usuaris. Es defineix un algoritme basat en deep learning o aprenentatge profund, que és una tècnica d'extracció i transformació de noves característiques del processament de la informació, les quals poden ser de forma supervisada o no. Són algoritmes que funcionen en un sistema per capes, simulant el funcionament bàsic del cervell que s'utilitza amb les neurones. És a dir, el conjunt de capes que forma el deep learning representen les neurones del cervell. Actualment la definició d'aprenentatge profund és un sinònim modern de les aplicacions de les xarxes neuronals.

L'aprenentatge profund es considera també una tècnica de machine learning o aprenentatge computacional que ensenya als computadors a fer el que és natural per als éssers humans, aprendre amb exemples. L'aprenentatge profund és una tecnologia clau en l'àmbit dels cotxes sense conductor, que els permet reconèixer una senyal de stop o distingir un vianant d'una farola. És també la clau del control per veu en els dispositius de consum com telèfons, televisors i altaveus de mans lliures. A l'aprenentatge profund, un model de computador aprèn a realitzar tasques de classificació directament a partir de imatges, text o so. Els models d'aprenentatge profund poden aconseguir una precisió inclús superior al rendiment humà. Els models s'entrenen fent servir un gran conjunt de dades etiquetades i arquitectures de xarxes neuronals que contenen múltiples capes. Els algoritmes d'aprenentatge profund contrasten amb els algoritmes d'aprenentatge poc profund pel número de transformacions aplicades a la senyal mentre es propaga des de la capa d'entrada fins a la capa de sortida. Cada una d'aquestes transformacions inclou paràmetres que es poden entrenar com pesos o llindars. No existeix un estàndard per al número de transformacions o capes que converteix a un algoritme en profund però la majoria de investigadors en el camp consideren que l'aprenentatge profund implica més de dos transformacions intermitjes. [23] Aquest algoritme consisteix en assignar una puntuació a cadascun dels estudis emmagatzemats a la base de dades i, cada cop que algun d'aquests estudis es seleccionat com a opció preferent per un usuari, aquesta punta-

ció s'incrementa en una unitat. El llistat d'estudis que proporciona l'assistent es mostra ordenat segons aquesta puntuació, amb prioritat sobre l'ordre establert per les notes de tall en el cas dels graus, de manera que el estudis que més es vagin triant es mostraran en una posició superior a altres estudis que els usuaris no trien mai. A continuació es mostra un diagrama de com es comuniquen el conjunt d'aplicacions que conformen el projecte.

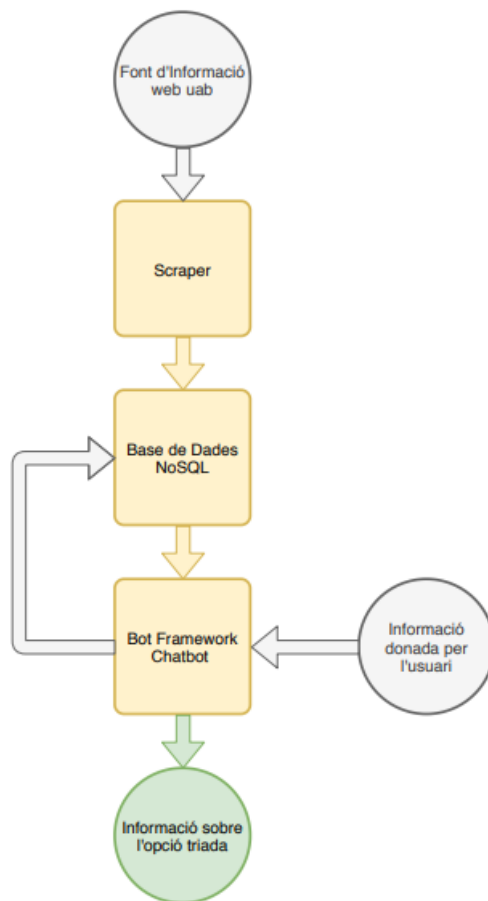


Fig. 4 Diagrama de l'aplicació

9 RESULTATS

Un dels objectius que s'ha aconseguit assolir de manera satisfactòria ha estat la obtenció de les dades a tractar des de la pàgina web de la UAB. La bona organització del codi de la web ha facilitat la utilització de les tècniques de scraping per seleccionar la informació necessària. En un principi es va plantejar l'opció de tractar les dades al vol dins del bot, però finalment s'ha fet servir la base de dades MongoDB, que al tractar-se d'una base de dades tipus NoSQL ha simplificat el filtrat i tractament de les dades, a més de dotar de persistència a aquestes dades. Fent servir l'eina MongoDB Compass, una interfície gràfica que permet analitzar el contingut de la base de dades, es pot comprovar fàcilment la informació obtinguda. En total s'ha obtingut informació sobre un total de 115 estudis de graus i 189 estudis de màsters oficials.

Collection Name ^	Documents	Avg. Document Size	Total Document Size	Num. Indexes	Total Index Size	Properties
choices	9	206.9 B	1.8 KB	1	36.0 KB	
degrees	115	1.1 KB	128.5 KB	1	36.0 KB	
masters	189	1.3 KB	251.1 KB	1	36.0 KB	

Fig. 5 Dades totals

S’ha aconseguit també desenvolupar un chatbot que consta d’un diàleg predefinit amb un conjunt de preguntes que permeten, a partir de les respostes indicades per l’usuari, incrementar la puntuació associada a la opció finalment escollida. D’aquesta manera les opcions mostrades en posteriors consultes d’altres usuaris s’adapten a la puntuació indicada per usuaris anteriors, completant així l’algorisme de deep learning plantejat.

Per tal d’avaluar el correcte funcionament de l’assistent virtual s’ha fet servir Bot Framework Emulator, que és una aplicació de escriptori que permet als desenvolupadors de bots provar i depurar els bots construïts fent servir Microsoft Bot Framework SDK. Es pot utilitzar aquest emulador per provar els bots que s’executen tant localment en la màquina de desenvolupament com els bots que s’executen de forma remota a través de un túnel.

A continuació es mostra una simulació d’execució on s’ha seleccionat estudis de grau de la branca d’enginyeries amb una nota de tall inferior a 7.

Escull una de les següents opcions seleccionades.

- Grau en Gestió de Ciutats Intel·ligents i Sostenibles Campus de Bellaterra [6.02]
- Grau en Enginyeria de Sistemes de Telecomunicació Campus de Bellaterra [5]
- Grau en Enginyeria Electrònica de Telecomunicació Campus de Bellaterra [5]
- Grau en Enginyeria Electrònica Industrial i Automàtica Campus de Barcelona [5]
- Grau en Enginyeria en Energies Renovables i Eficiència Energètica Barcelona [5]
- Grau en Enginyeria en Organització Industrial Campus de Barcelona [5]
- Grau en Enginyeria Mecànica Campus de Barcelona [5]
- Grau en Informàtica i Serveis Campus de Sant Cugat del Vallès [5]

Fig. 6 Simulació d’execució selecció d’opcions

S’ha triat la opció de Grau en Enginyeria Mecànica. A continuació es mostra la resposta indicada per l’assistent.

Grau en Enginyeria Mecànica

Campus de Barcelona

Nota de tall: 5

Esdevindràs l'enginyer o enginyera que resol els reptes de l'enginyeria dissenyant el millor producte, escollint els materials adequats per a cada aplicació industrial i planificant com es fabrica el producte.

[Més informació](#)

Fig. 7 Simulació d’execució resposta obtinguda

Després de realitzar aquesta selecció s’ha incrementat la puntuació associada al Grau en Enginyeria Mecànica en una unitat, de manera que si un usuari torna a realitzar la mateixa cerca que l’usuari anterior veurà com aquesta opció té prioritat respecte a les altres opcions que no han estat triades per cap altre usuari. A continuació es mostra el llistat d’opcions disponibles corresponents a la mateixa selecció anterior de estudis de grau de la branca d’enginyeries amb una nota de tall inferior a 7. Es pot comprovar que l’ordre en el que es mostren les opcions disponibles s’ha modificat tenint en compte que, en aquest cas, a diferència del cas anterior, el Grau en Enginyeria Mecànica té una puntuació superior a la resta d’opcions disponibles.

Escull una de les següents opcions seleccionades.

- Grau en Enginyeria Mecànica Campus de Barcelona [5]
- Grau en Gestió de Ciutats Intel·ligents i Sostenibles Campus de Bellaterra [6.02]
- Grau en Enginyeria de Sistemes de Telecomunicació Campus de Bellaterra [5]
- Grau en Enginyeria Electrònica de Telecomunicació Campus de Bellaterra [5]
- Grau en Enginyeria Electrònica Industrial i Automàtica Campus de Barcelona [5]
- Grau en Enginyeria en Energies Renovables i Eficiència Energètica Barcelona [5]
- Grau en Enginyeria en Organització Industrial Campus de Barcelona [5]
- Grau en Informàtica i Serveis Campus de Sant Cugat del Vallès [5]

Fig. 8 Simulació d’execució selecció d’opcions amb prioritat

10 CONCLUSIONS

En aquest article s'ha plantejat en primer lloc el problema a resoldre dins l'àmbit de la demanda d'informació referent als estudis oferts per un centre educatiu, en aquest cas la UAB. A continuació s'han detallat els objectius a assolir durant la realització del projecte, que són la tria de l'entorn de treball a utilitzar, definir la font d'informació de les dades a tractar i com emmagatzemar aquestes dades, definir el conjunt de preguntes que l'assistent realitzarà a l'usuari i, finalment definir el mètode per a tenir en compte les tries finals realitzades per l'usuari. S'ha realitzat un estudi sobre l'estat de l'art en l'àmbit dels chatbots i s'ha detallat la metodologia utilitzada per tal de dur a terme el desenvolupament. El fet de que la UAB no disposi d'una api d'accés públic a la informació proporcionada en el seu web ha suposat una dificultat afegida a l'hora de realitzar el desenvolupament d'aquest projecte. A nivell personal es considera que ha estat un projecte molt interessant que ha permès la immersió en el món dels chatbots, un camp àmpliament utilitzat avui en dia i amb moltíssima projecció de futur. Donada la gran quantitat d'informació i exemples disponibles sobre el desenvolupament de chatbots fent servir Microsoft Bot Framework, es recomana aquest projecte com a punt de partida a l'hora d'abordar la utilització de chatbots en l'àmbit dels centres educatius.

11 LÍNIES DE TREBALL FUTURES

Es plantegen els següents temes a tractar com a possibles línies de millora sobre el desenvolupament realitzat.

- Extreure estadístiques d'ús sobre la informació aportada pels usuaris, com per exemple estudis més demandats o notes d'accés més populars.
- Integració del chatbot amb aplicacions de missatgeria instantània.
- Dotar al chatbot de la capacitat de reconeixement de llenguatge natural.

AGRAÏMENTS

Agrair al meu tutor Joan Giner que des del primer moment ha estat molt atent al seguiment del treball i sempre disposat a adaptar-se en els horaris.

Agrair també a la meva tia Carmen, sense la seva insistència no hagués recorregut aquest tram del camí.

BIBLIOGRAFIA

- [1] Viquipèdia. Bot conversacional. Ago. 2019. [Online]. Disponible: https://ca.wikipedia.org/wiki/Bot_conversacional
- [2] Wikipèdia. Chatbot. Feb. 2020. [Online]. Disponible: <https://en.wikipedia.org/wiki/Chatbot>
- [3] Rainer Winkler. Matthias Söllner. Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis. Mar. 2018. Conference Paper.
- [4] Dan Shewan. 10 of the Most Innovative Chatbots on the Web. Feb. 2020. [Online]. Disponible: <https://www.wordstream.com/blog/ws/2017/10/04/chatbots>
- [5] Michael Keenan. The 14 Best Chatbot Examples in 2020 (and How to Build Your Own). Gen. 2020. [Online]. Disponible: <https://manychat.com/blog/chatbot-examples>
- [6] Ester Chicano Tejada. Chatbot: La inteligencia artificial en las aplicaciones de mensajería. Set. 2019. [Online]. Disponible:

<https://revistadigital.inesem.es/gestion-empresarial/que-es-un-chatbot>

[7] Andrés Pedreño Muñoz. Lola, el chatbot inteligente que triunfa entre los estudiantes. Des. 2018. [Online]. Disponible: https://retina.elpais.com/retina/2018/11/30/innovacion/1543580663_865121.html

[8] Universidad de Murcia. La Universidad de Murcia presenta a LOLA, un asistente de inteligencia artificial para ayudar a los nuevos alumnos. Jul. 2018. [Online]. Disponible: <https://www.um.es/web/sala-prensa/-/la-universidad-de-murcia-presenta-a-lola-un-asistente-de-inteligencia-artificial-paraayudar-a-los-nuevos-alumnos>

[9] Wikipèdia. Dialogflow. Jul. 2019. [Online]. Disponible: <https://en.wikipedia.org/wiki/Dialogflow>

[10] [Online]. Disponible: <https://discover.bot/bot-talk/beginners-guide-bots/popular-chatbot-frameworks>

[11] Universitat Autònoma de Barcelona. Dipòsit Digital de Documents de la UAB. Abr. 2020. [Online]. Disponible: <https://www.uab.cat/web/els-nostresfons/diposit-digital-de-documents-de-la-uab-1345754055775.html>

[12] Wikipèdia. Web scraping. Mar. 2020. [Online]. Disponible: https://en.wikipedia.org/wiki/Web_scraping

[13] Wikipèdia. Node.js. Mar. 2020. [Online]. Disponible: <https://en.wikipedia.org/wiki/Node.js>

[14] Grohs Fabian. Learn Web Scraping with NodeJs in 2020 - The Crash Course. Mai. 2019. [Online]. Disponible: <https://www.udemy.com/course/nodejsweb-scraping/>

[15] Grohs Fabian. 4 Easy Steps to Web Scraping with NodeJs. Nov. 2018. [Online]. Disponible: <https://learnscraping.com/4-easy-steps-to-web-scraping-with-nodejs/>

[16] Node.js Foundation. Express. Abr. 2020. [Online]. Disponible: <https://expressjs.com/>

[17] Wikipèdia. MongoDB. Jun. 2020. [Online]. Disponible: <https://en.wikipedia.org/wiki/MongoDB>

[18] Mozilla. async function. Mai. 2020. [Online]. Disponible: https://developer.mozilla.org/en-US/docs/Web/JavaScript/Reference/Statements/async_function

[19] Anton Lavrenov. JavaScript loops - how to handle async/await. Oct. 2017. [Online]. Disponible: <https://lavrton.com/javascript-loops-how-to-handle-async-await-6252dd3c795/>

[20] npm. generator-botbuilder. Mai. 2020. [Online]. Disponible: <https://www.npmjs.com/package/generator-botbuilder>

[21] Microsoft. How bots work. Gen. 2020. [Online]. Disponible: <https://docs.microsoft.com/en-us/azure/bot-service/bot-builder-basics?view=azure-bot-service-4.0&tabs=javascript>

[22] Microsoft. Dialogs library. Mai. 2020. [Online]. Disponible: <https://docs.microsoft.com/en-us/azure/bot-service/bot-builder-concept-dialog?view=azure-bot-service-4.0>

[23] The MathWorks, Inc. What Is Deep Learning? Gen. 2020. [Online]. Disponible: <https://www.mathworks.com/discovery/deep-learning.html>

