

# PLA DE CARRERA

Mireia Gàlvez Castro

Enginyeria Tècnica en Informàtica de Gestió  
Escola Politècnica d'Enginyeria de Vilanova i la Geltrú  
Universitat Politècnica de Catalunya

## 1.- Introducció

Avui, a Espanya, més de 6 milions de persones, que representen més del 27% de la població en edat de treballar es troben en situació d'atur i les previsions no conviden a l'optimisme a curt termini. Aquestes persones en situació d'atur s'han de posar davant d'un ordinador per cercar treball i per poder tenir més opcions volen saber quins són els llocs de treball on encara es demana gent, que poden estudiar per trobar treball, ... Per donar resposta a aquestes preguntes és útil una eina que ens informi de totes aquestes dades.

Actualment, a Espanya no existeixen serveis d'informació i coneixement en l'àmbit laboral com el que descrivim.

Si que existeixen, alguns agregadors de feina internacionals, serveis que tenen certes similituds amb els descrits anteriorment com ara "SimplyHired.com" o "Indeed.com", però es centren principalment en descriure les tendències de publicació d'ofertes de treball en el mercat en el que operen.

El motiu principal pel que la informació d'aquests "sites" internacionals no es útil pels usuaris d'Espanya és perquè la informació que mostren no es

representativa del mercat laboral al que pertanyen (Espanya).

També existeix un portal ("jobtitled.com") que té una idea similar i que es basa en el perfil dels seus usuaris a la xarxa professional "LinkedIn". No obstant, el volum de dades és baix i la informació que hi apareix correspon principalment al mercat anglosaxó.

Per tant, podem concloure que aquest projecte és pioner al nostre país en un terreny verge i ampli per córrer.

Aquest projecte es desenvolupa juntament amb l'empresa Infojobs.

La orientació d'aquest projecte es basa en obtenir informació dels currículums dels candidats del portal de feina d'Infojobs.

Per aquest projecte l'objecte essencial és el currículum del candidat. Aquests currículum ens mostren el conjunt d'experiències a llarg d'una trajectòria professional, d'aquesta manera ens pot ajudar a descobrir patrons de comportament que puguin ser posats a l'abast dels usuaris, en forma d'informació i coneixement. Pot ser una eina útil per orientar la seva trajectòria

professional i incrementar les seves opcions d'ocupació.

## 2.- Objectiu

L'objectiu del projecte és desenvolupar una aplicació web interactiva que ofereixi informació per ajudar a qualsevol persona a orientar la seva carrera professional en base a la informació i patrons de comportament recollit en els currículums dels usuaris registrats al portal de feina Infojobs.

Els apartats que ens interessin són les experiències desenvolupades i els estudis realitzats durant la trajectòria professional de cada persona. La combinació d'aquests ens descriuen la seva trajectòria professional, és a dir, quins passos han dut a terme al llarg de la seva vida per arribar fins al punt present.

Per tot això, s'estableix com hipòtesi en aquest projecte, que dins del conjunt de dades d'InfoJobs hi ha trajectòries professionals que comparteixin elements comuns i que, per tant, sorgeixen patrons de comportament que siguin fonts de coneixement per ser posats a l'abast de la comunitat com a eina d'orientació i planificació de la carrera professional.

Aquesta informació ha de ser utilitzada principalment com a eina de referència i suport a l'hora de prendre decisions en base a la comparació d'un mateix respecte a la resta de la comunitat.

El resultat del projecte proposat serà un portal web on es mostraran gràfiques que descriuran els patrons de comportament de cada lloc de treball a nivell estadístic, fent una explotació intensiva de les relacions entre els elements d'informació, és a dir, es mostraran gràfiques que

descriuen els patrons de comportament dels llocs de treball, dels centres de formació, de les empreses i els estudis a nivell estadístic.

## 3.- Tecnologies

Ja que el projecte es desenvolupa juntament amb l'empresa InfoJobs hi ha cert condicionament en algunes de les eines i tecnologies a utilitzar. Algunes d'aquestes són:

- . "Highcharts": llibreria que permet la representació i la càrrega de gràfics de forma dinàmica mitjançant "jQuery". Aquesta característica facilitarà la interacció de l'usuari amb l'aplicació a l'hora d'aplicar criteris de filtratge i refinament de forma individual sobre una única gràfica. S'ha escollit aquesta llibreria perquè és compatible amb HTML5 (per posterior implementació a mòbil).

- . Aplicació web: s'ha utilitzat com llenguatge Java i el template Spring.

- . Base de dades de graf: Neo4J (base de dades no relacional). S'ha tingut en compte perquè actualment té una presència creixent en el mercat, és desenvolupada amb Java, és transaccional i disposa d'una API que permet la creació de recorreguts transversals.

- . Neteja de dades: no hem trobat cap eina que s'ajusti a les necessitats del projecte. Per això és part del projecte. Però si que hem trobat inspiració en l'eina "Google Refine" per poder agrupar, transformar i netejar les dades.

. "Memcache": sistema de caché d'objectes de memòria per accedir ràpidament als resultats emmagatzemats. D'aquesta manera no realitzem la mateixa consulta varies vegades, sinó que se serveix la informació ja obtinguda anteriorment, solament es fa una nova consulta a Neo4J si els paràmetres canvien perquè és una nova consulta.

. BBDD SQL: sistema de seguretat per si "memcahe" no estigués operatiu. S'ha creat una taula on es guarden les consultes realitzades i solament es va a buscar la consulta a Neo4J si té paràmetres nous i així s'evita consultes costoses.

#### 4.- Procés

El factor clau és la informació que extraïem dels camps que trobem en la secció d'experiències dels currículums d'Infojobs: professió (lloc de treball), nom d'empresa, centre de formació i estudis. En aquests camps es dóna l'opció d'introduir valors lliures. Això provoca que l'informació introduïda no sigui fiable ja que tenim valors semànticament equivalents i valors similars escrits. Això provoca que s'hagi de fer inicialment un estudi per determinar la viabilitat del projecte. S'ha tingut en compte:

. Volum de dades suficient per extreure patrons de comportament comuns. En aquest projecte es treballarà amb els candidats que s'han donat d'alta en Infojobs els últims 48 mesos, és a dir, amb una mostra de:

- o 4.909.265 currículums,
- o 3.930.848 candidats amb experiències,
- o 4.142.929 candidats amb estudis.

. Qualitat i consistència de la informació. Un estudi inicial ha permès veure que la qualitat de la informació és satisfactòria per la realització d'aquest projecte, si bé es detecten registres d'informació no vàlids, com per exemple experiències on la data de finalització és anterior a la data d'inici, candidats sense data de naixement, ... Els registres que no tenen informació vàlida i consistent es descarten per tal que no contaminin la informació vàlida.

. Normalització del camps de lloc de treball, empresa i centre de formació. Aquests camps tenen alguns valors que no són correctes, per això hem normalitzat, que es basa en:

- o conversió a majúscules,
- o eliminació de signes de puntuació (accents, dièresis, punts, comes, ...),
- o substitució de caràcters especials (ñ, °, ^, ),
- o eliminació d'espais inicials i finals,
- o eliminació de números,
- o eliminació de "stopwords" (articles, preposicions, ...).
- o substitució de sinònims.
- o aplicació d'expressions regulars per convertir cadenes senceres en valors normalitzats i que actuïn d'agrupadors de valors.

. Procés de càrrega del graf utilitzant el fitxers de dades que han estat prèviament tractats pel procés de normalització. És en aquest punt quan s'ha de precalcular alguna de la informació que després s'utilitzarà a l'aplicació, amb l'objectiu d'optimitzar el temps de càlcul.

. Presentació de resultats. És on es visualitzen els resultats mitjançant l'aplicació web interactiva desenvolupada amb Java.

## 5.- Conclusions

Fonamentar el projecte sobre camps de dades (lloc de treball, empresa, centre de formació i estudis) ha estat clau per poder extreure'n informació.

L'estudi inicial de les dades, previ a la normalització, ha fet aflorar incoherències en alguns camps d'informació motivats per errors o per necessitats no cobertes per InfoJobs, com ara candidats amb estudis universitaris o màsters que no indiquen els seus estudis anteriors.

Per aconseguir generar coneixement útil s'ha hagut de passar pel peatge d'haver de construir eines auxiliars per normalitzar les dades i millorar-ne la qualitat.

Hi ha múltiples opcions per resoldre el repte d'aquest projecte, però la utilització d'un graf com a mecanisme de suport de les dades ha simplificat i reduït el temps de càlcul de les gràfiques que presenta l'aplicació.

### 5.1- El futur

Aquest projecte s'ha convertit en un projecte d'empresa que serà el referent sobre el que es fonamentarà el nou InfoJobs que pretén proporcionar informació de valor i individualitzada a cada usuari al llarg de la seva carrera professional. Algunes de les funcionalitats que s'estan desenvolupant sobre InfoJobs tindran un vincle amb aquesta aplicació.

Fins ara, l'eina és informativa i passiva, i no ajuda a fer el seguiment dels objectius professionals que s'hagi pogut posar l'usuari. Per fer això cal desenvolupar una

part privada, que a partir de conèixer l'usuari, ens permeti establir un diàleg amb ell per tal de fixar objectius professionals i terminis, fer-ne el seguiment, així com suggerir-li camins per arribar-hi (ofertes laborals, formació i xarxa professional).

Aquest projecte està impulsant un canvi tecnològic a InfoJobs que possibilitarà el desenvolupament de noves funcionalitats de valor pels usuaris que fins ara no eren possibles. Algunes d'aquestes funcionalitats fan referència a vestir l'apartat de xarxa de relacions professional orientada a la cerca de treball.