

Eina per a la construcció d'arbres genealògics

Martí Saladelafont Díaz

Resum– Aquest projecte presenta una eina que permet trobar l'arbre genealògic d'una persona automàticament a partir d'una base de dades. La base de dades utilitzada correspon als matrimonis efectuats a la diòcesi de Barcelona entre els segles XVI i XX, documentació ubicada als llibres d'Esposalles de la Catedral. El programa analitza la base de dades i, a partir de la selecció d'una persona, busca el seu arbre genealògic. Aquest programa permet veure gràficament l'arbre generat, interactuar amb ell i també permet generar un arxiu PDF amb les dades de cada un dels avantpassats, ordenats segons l'estructura de Sosa-Stradonitz. Aquesta eina ha permès obtenir de manera satisfactòria els arbres genealògics que s'han buscat, atesa la dificultat de trobar l'arbre genealògic autèntic de la manera tradicional.

Paraules clau– Genealogia, graf, arbre genealògic, Sosa-Stradonitz, base de dades, Breadth First Search (BFS).

Abstract– This project presents a tool that allows to find the genealogical tree of a person automatically from a data base. The database used for this project corresponds to the marriages effected to the diocese of Barcelona between the 16th and the 20th centuries, documentation stored in the Esposalles' books about the Cathedral. The program analyses the database and, as of the selection of a person, looks for his genealogical tree. This program allows to see graphically the tree, interact with it, and also allow to generate a PDF document with the data of each one of the ancestors of the person, ordered as the Sosa-Stradonitz's structure. This tool has allowed to obtain of satisfactory way the genealogical trees that have looked for, due to the difficulty to find the authentic genealogical tree by the traditional way.

Keywords– Genealogy, graph, genealogical tree, Sosa-Stradonitz, data base, Breadth First Search (BFS).

1 INTRODUCCIÓ

LA genealogia és la ciència que estudia l'ascendència de les persones i els seus llinatges, i n'estudia l'origen, descendència i aliances. Ens pot permetre la recerca i estudis en diferents camps de la sociologia, com poden ser la recerca demogràfica o la història.

Moltes de les dades històriques que podem utilitzar per a estudiar l'ascendència d'una persona les podem trobar en arxius eclesiàstics, notariais o municipals. Els grans volums de dades que tenen aquests diferents arxius fan que buscar la

genealogia d'una persona sigui una tasca prou difícil. Són de molt ajut programes o webs que permeten tenir un control sobre la cerca a realitzar. Alguns exemples poden ser FamilySearch [6], Ancestry [1] o MyHeritage [8]. També són d'ajuda buidatges realitzats per persones de part d'aquests arxius, com els que es poden trobar a la Societat Catalana de Genealogia [3].

Aquest treball està orientat als arxius eclesiàstics que ens podem trobar a les diferents parròquies. Concretament, s'utilitza una base de dades a partir de la informació extreta dels llibres de matrimonis registrats a la Catedral de Barcelona.

La base de dades d'ESPOSALLES és una recopilació de les dades extretes dels Llibres d'Esposalles, conservats a l'arxiu de la Catedral de Barcelona. Aquesta col·lecció de llibres conté la informació d'aproximadament 600,000 matrimonis registrats del voltant de 250 parròquies de la

-
- E-mail de contacte: mrtinet92@gmail.com
 - Menció realitzada: Computació
 - Treball tutoritzat per: Josep Lladós Canet (Ciències de la Computació)
 - Curs 2016/17

diòcesi de Barcelona, compresos entre els anys 1451 i 1905.

Cada una d'aquestes llicències de matrimonis registrades als llibres d'Esposalles contenen la informació que correspon al matrimoni, com pot ser el nom i cognom dels contractants, el nom i cognom dels pares, la professió o el lloc de residència, a part del lloc on se celebra l'ofici.

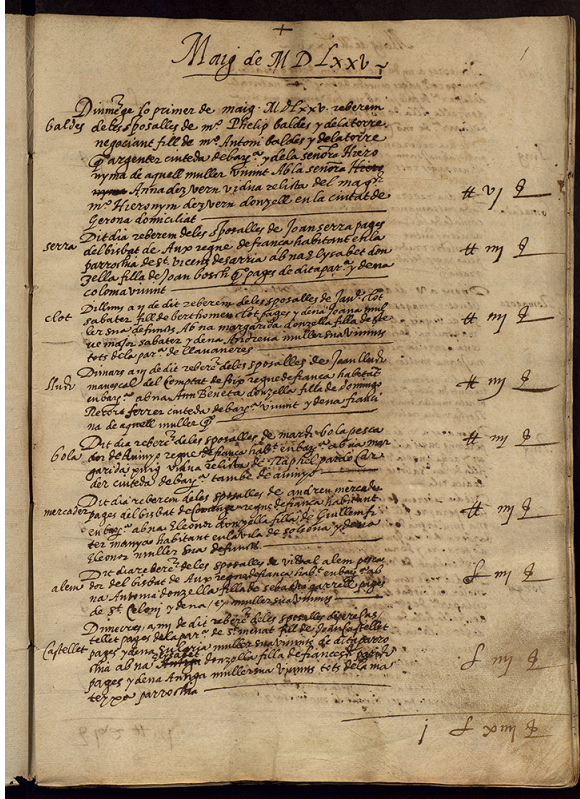


Fig. 1: Exemple dels Llibres d'Esposalles

1.1 Motivació

La meua motivació per a la realització d'aquest treball rau en el fet del meu gran interès en la història. Des de petit que sempre m'ha interessat conèixer el passat, de la història en general i més concretament el passat de la meua família. A causa d'aquest interès, fa ja uns anys que vaig començar la tasca de fer el meu arbre genealògic, visitant diversos arxius, tant parroquials com civils. Amb això m'he adonat com és de complicat la realització d'aquest arbre, ja que sovint tens molta més informació del que és realment necessària, fet que comporta una gran despesa de temps buscant en els diferents llibres i arxius.

Així doncs, vaig trobar encertat elaborar una eina que et permetés fer més àgil aquesta cerca i que pogués ser útil a més gent.

1.2 Objectius

L'objectiu principal d'aquest treball és la creació d'una eina intuïtiva que faciliti la connexió entre les diferents dades que conté la base de dades i ens pugui mostrar l'arbre genealògic d'una persona concreta.

Aquestes connexions poden ser de diferent tipus, ja sigui per raons familiars (pares i fills), per raons de procedència (població de naixement) o laborals (mateixa ocupació).

Com a subobjectius que té aquest treball podem trobar els següents:

- Construir enllaços genealògics entre els diferents registres.
- Implementar algorismes de seguiment de camins en grafs.
- Implementar una aplicació de cerca de registres

1.3 Metodologia

Atès que aquest treball es realitza de manera individual, la metodologia que s'aplicarà per a la realització del treball serà una metodologia iterativa, inspirant-se amb metodologies àgils com pot ser SCRUM.

En aquest cas se seguirà l'esquema de *sprints*, on cada *sprint* serà el període entre els diferents lliuraments del treball, i al final de cada un hi haurà un informe de seguiment del progrés realitzat.

Tenint en compte aquesta premissa, es realitzaran reunions periòdiques amb el tutor, que en aquest cas tindrà el rol de client. Utilitzar aquesta metodologia permetrà desenvolupar les tasques a realitzar tot assumint els possibles canvis que puguin sorgir al llarg del projecte, i d'aquesta manera incorporar-los al projecte sense inconvenients.

TAULA 1: TAULA DE PLANIFICACIÓ

1. Estudi Bibliogràfic	48 hores	12.02	25.02
1.1 Consultar estat de l'art	12 hores		
1.2 Consultar bibliografia	12 hores		
1.3 Revisar la base de dades	24 hores		
2. Disseny d'algorismes	10 hores	25.02	11.03
2.1 Consultar documentació	4 hores		
2.2 Pseudo-implementació	6 hores		
3. Implementació dels algorismes	105 hores	11.03	21.04
3.1 Consultar documentació	5 hores		
3.2 Implementació	100 hores		
4. Implementació de l'aplicació	105 hores	21.04	21.05
4.1 Consultar documentació	2.5 hores		
4.2 Implementació	100 hores		
4.3 Proves	2.5 hores		
5. Preparació del treball final	32 hores	21.05	29.06
5.1 Preparació article final	21 hores		
5.2 Preparació de la presentació	11 hores		

1.4 Planificació

A partir dels objectius prèviament definits, al cronograma que ens mostra la taula 1 s'hi ha detallat les tasques i subtasques a realitzar i la seva durada.

En ell s'hi destaquen cinc tasques principals: Estudi Bibliogràfic, Disseny d'Algorismes, Implementació dels Algorismes, Implementació de l'Aplicació i Preparació del Treball Final.

A la secció d'Apèndix (8) hi ha el diagrama de Gantt seguit per a la realització del treball.

2 ESTAT DE L'ART

Actualment existeixen diverses aplicacions per a la realització d'arbres genealògics. Des d'aplicacions web gratuïtes, com poden ser Geneanet [7], MyHeritage [8] o Ancestry [1], com també aplicacions de software, ja siguin gratuïtes (Gramps [11]) o de pagament (GDS [2]).

Aquestes aplicacions permeten mantenir un control sobre tota la documentació i informació relacionada amb un arbre genealògic. Moltes aplicacions també permeten veure l'arbre d'una forma gràfica que facilita la comprensió de totes les dades que té la base de dades, i permeten una navegació interactiva amb l'arbre.

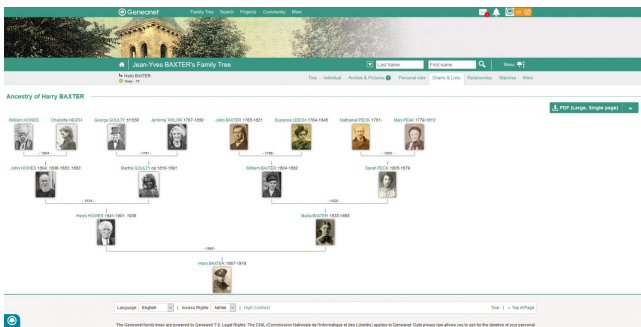


Fig. 2: Exemple de la pàgina web Geneanet

Malgrat això, la gran majoria d'aquestes aplicacions únicament permeten la construcció de l'arbre a mesura que es van introduint les persones a la base de dades. Així doncs, anar completant l'arbre és una tasca molt laboriosa.

El problema d'aquestes aplicacions en comparació amb el problema de l'aplicació plantejada en aquest treball és el fet que, en aquest treball, s'utilitzarà una base de dades molt gran i a partir d'aquesta s'extraurà automàticament un arbre genealògic. Les aplicacions que s'han comentat no permeten realitzar aquesta tasca, i per tant no serien útils per al cas que ens ocupa.

3 BASES DE DADES

El primer que s'ha realitzat per tal de començar aquest treball ha estat la recopilació de les dades necessàries per crear la base de dades de prova, que ha servit per a la realització dels algorismes.

Abans de començar amb la base de dades d'Esposalles, s'ha decidit fer servir dues bases de dades diferents. En primer lloc he utilitzat la base de dades dels matrimonis que tinc recopilats de la recerca genealògica de la meua família. Això es deu al fet que, en saber ja qui és l'arbre correcte

amb les seves connexions, ens permet fer una comprovació ràpida de la validesa dels algorismes que s'han de dissenyar.

En segon lloc s'ha utilitzat una base de dades dels matrimonis de Sant Llorenç d'Hortons. Aquesta base de dades s'ha extret d'un buidatge dels llibres de matrimonis que es conserven a la parròquia de la localitat. És una base de dades pública que es pot trobar a la web de la Societat Catalana de Genealogia. Aquesta base de dades és molt més completa que l'anterior, i el fet de no saber quins són els arbres concrets ni les connexions entre les diferents persones fa que puguem comprovar si els algorismes funcionen correctament.

Un cop obtingudes les bases de dades de prova, s'ha començat a utilitzar la base de dades d'Esposalles. Primerament s'ha utilitzat únicament una part de la base de dades per tal de poder realitzar les proves corresponents.

Per finalitzar, s'ha creat una base de dades a partir del graf generat. Això és degut al fet que a més volum de dades que contingui la base de dades d'on extraiem la informació per a la construcció del graf més temps ens tarda a realitzar-lo. Així doncs, per tal que en futures consultes el temps en carregar sigui més reduït.

3.1 Base de dades familiar

La base de dades que s'ha utilitzat per poder començar a implementar l'aplicació és la que es correspon a les dades que he anat recopilant al llarg dels anys sobre la meua família. Comprèn 36 matrimonis des del segle XVII fins al segle XX.

La base de dades té dues taules. A la primera, anomenada *matrimonis*, hi ha tota la informació referida a cada un dels matrimonis, com ara la data, la població i les dades de cada un dels contraents. En segon lloc hi ha la taula *persones*, que conté la informació de cada una de les persones de la base de dades. La figura 3 ens mostra l'estructura.

matrimonis	persones
id INT(11)	id INT(11)
data_m atrim oni VARCHAR(90)	id_m atrim oni INT(11)
nom_m arit VARCHAR(90)	nom VARCHAR(90)
cognom_1_m arit VARCHAR(90)	cognom 1 VARCHAR(90)
cognom_2_m arit VARCHAR(90)	cognom 2 VARCHAR(90)
nom_pare_marit VARCHAR(90)	nom_pare VARCHAR(90)
nom_m are_marit VARCHAR(90)	nom_m are VARCHAR(90)
nom_m uller VARCHAR(90)	genera VARCHAR(90)
cognom_1_m uller VARCHAR(90)	job VARCHAR(90)
cognom_2_m uller VARCHAR(90)	
nom_pare_muller VARCHAR(90)	
nom_m are_muller VARCHAR(90)	
poblacio VARCHAR(90)	
Indexes	Indexes

Fig. 3: Estructura de la base de dades familiar

Cal constatar que cada un dels noms i dels cognoms registrats a la base de dades han estat normalitzats.

3.2 Base de dades de Sant Llorenç d'Hortons

Aquesta base de dades, a diferència de l'anterior, ja és una base de dades més extensa. Correspon al buidatge dels llibres de matrimonis que es conserven a la parròquia de Sant Llorenç d'Hortons (Alt Penedès) [4]. Aquest buidatge conté els registres des del segle XVI fins a principis del segle XX.

La seva estructura és la mateixa que a la base de dades familiar. Consta de 1485 matrimonis i, en aquest cas, s'ha mantingut l'ortografia original de la transcripció.

3.3 Base de dades d'Esposalles

La base de dades d'Esposalles és la principal base de dades en què es basa aquest treball. Correspon a les dades recopilades als llibres d'Esposalles de la Catedral de Barcelona. Comprenen al voltant de 600,000 matrimonis al llarg de cinc segles, des del segle XV fins a principis del XX.

La base de dades consta de dues taules, la taula *persones* i la taula *transcripcions*. A la taula *persones* hi consten els registres de cada un dels integrants del matrimoni (marit i muller), així com els pares dels dos contraents, en el cas que constin als documents. A la taula *transcripcions* hi consta cada un dels matrimonis registrats als documents, amb la informació de tots els que hi prenen part.

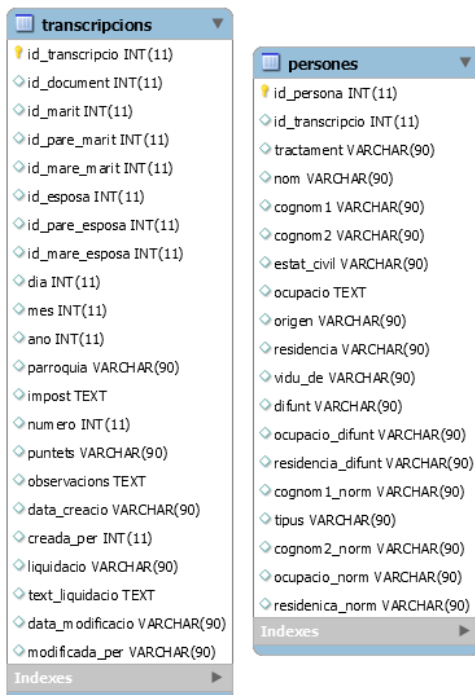


Fig. 4: Estructura de la base de dades d'Esposalles

Per culpa del fet que la base de dades és molt extensa, i la falta de permisos per poder-la extreure del Centre de Visió per Computador [5], s'ha hagut de realitzar les proves d'aquesta base de dades amb una petita part d'aquesta. Aquesta part constava d'uns 30,000 registres a la taula de *persones*, i uns 5,000 a la taula de *transcripcions*, i

comprenia matrimonis entre el 1575 i el 1619.

S'ha de tenir en compte que molta de la informació guardada en aquesta base de dades no és necessària per a la realització del còmput dels possibles pares d'una persona.

3.4 Base de dades del graf generat

Un cop s'ha generat el graf de relacions de les bases de dades, s'ha creat una nova taula a la base de dades que s'utilitza per a guardar les diferents connexions que hi ha.

Aquesta taula conté una relació entre l'identificador del node fill i l'identificador del pare o mare, en el cas que existeixi una relació. En el cas que no existeixi cap relació, és a dir, que el node no tingui pares, únicament hi consta el registre de l'identificador del node fill.

Així cada cop que s'utilitza la base de dades no s'ha de generar el graf de nou, sinó que es carrega a partir d'aquesta taula. Això ens permet obtenir el graf d'una forma molt més ràpida.

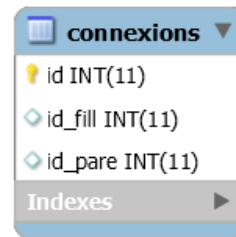


Fig. 5: Estructura de la taula connexions

3.5 Eines per a la base de dades

La base de dades s'ha creat en una plataforma MySQL [9]. Això ha permès que, al ser compatible amb el llenguatge SQL, sigui fàcilment exportable i es pugui utilitzar en diversos entorns. Cada una de les taules creades en les diferents bases de dades són compatibles en aquest format.

El programa utilitza aquesta plataforma per poder accedir a les bases de dades.

4 IMPLEMENTACIÓ

En la implementació d'aquest treball s'hi poden veure dues parts diferenciades. La primera part consisteix en la creació dels algorismes necessaris per a generar un graf de connexions a partir de la base de dades que s'hi indiqui i de realitzar les diferents cerques que realitzarem per obtenir l'arbre genealògic.

La segona part és la implementació d'un entorn gràfic que permeti fer més intuïtiu la recerca dels avantpassats i que ens faciliti la interacció amb el graf, a més que ens proporcioni la informació requerida de forma més amigable.

5 IMPLEMENTACIÓ DELS GRAFS

En la implementació dels grafs que s'han de realitzar en aquest treball podem diferenciar dues parts. La primera consisteix en la creació del graf de connexions, on tots els nodes (en aquest cas, persones) estan connectats amb els seus nodes pare.

La segona part consisteix en la creació del graf de l'arbre genealògic a partir d'una persona seleccionada.

5.1 Creació del graf de connexions

El primer pas de la implementació dels diferents algorismes ha estat l'elecció del llenguatge amb què s'ha realitzat l'aplicació. En aquest cas s'ha escollit el llenguatge Python [12], ja que és un llenguatge àgil i fàcil de programar. A més, aquest llenguatge té una llibreria anomenada NetworkX [10] que ens ajuda en la creació de grafs i amb la interacció amb ells.

Un cop escollit el llenguatge de programació, i amb les dades obtingudes de la base de dades, s'ha procedit a començar la implementació de l'algorisme.

Per a treballar correctament amb les dades obtingudes, s'han creat dues classes de Python:

- Classe Person: correspon a cada una de les persones que intervenen en un matrimoni, és a dir, el marit i la muller. Aquesta classe guarda el nom, els cognoms, l'ocupació, el gènere (masculí o femení), el lloc de residència, el lloc d'origen i l'identificador del matrimoni on participa. També té un identificador únic per a cada persona de la base de dades.
- Classe Marriage: correspon a cada un dels matrimonis dels quals es compon la base de dades. Conté la data del casament, el lloc on es va realitzar i un identificador únic.

S'ha recorregut cada entrada de la base de dades, és a dir de les persones que estan registrades a cada un dels matrimonis, i s'ha anat afegint com a node en el graf de relacions.

S'ha de puntualitzar que les persones que s'han afegit al graf són únicament el marit i la muller de cada matrimoni. Això es deu al fet que, encara que a la base de dades, per a cada matrimoni, s'hi inclogui també els pares dels dos contraents, incloure'ls al graf suposaria la duplicació de persones, ja que a la base de dades no es té en compte si els progenitors ja hi estan inscrits.

Després d'afegir tots els nodes al graf es comença a buscar les possibles relacions entre ells. De manera bàsica, les relacions que es busquen són de pares i fills, per tal d'obtenir correctament l'arbre genealògic. Per a la realització d'aquesta tasca es calcula la probabilitat que un node sigui el pare de l'altre. Com que en els registres històrics es premiava més la informació del marit en comptes de la muller, s'ha tingut en compte els homes per tal de calcular-ho.

Per al càlcul de la probabilitat s'ha tingut en compte la similitud dels cognoms, la coincidència en la població de casament, la població d'origen (comparant-la amb la població de residència del pare), la feina i que el període de temps comprès entre els dos matrimonis sigui inferior als cinquanta anys i superior als vint. Donant diferents pesos als diferents paràmetres es calcula una probabilitat que es correspon a la probabilitat que el node sigui el pare. Per tal d'assolir la màxima seguretat que el pare és el correcte, s'ha considerat que la probabilitat ha de ser més gran d'un 65%. Per sota d'aquest percentatge, es considera que el node no té cap pare.

Amb això obtenim una llista de possibles pares del node, ordenada de major probabilitat a menor.

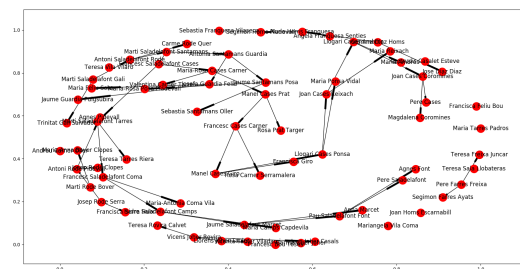


Fig. 6: Graf de connexions amb la base de dades familiar

A partir de la llista de possibles pares, es busca la muller del matrimoni on està registrat el possible pare. Es comprova si aquesta dona es correspon correctament amb la mare del node. Per fer-ho, es calcula el tant per cent de similitud entre el segon cognom del node i el primer de la dona (en el cas que hi sigui), com també la similitud entre la localitat de residència de la dona amb la de naixement del node. Donant pesos diferents a cada un d'aquests valors, es calcula la probabilitat que la dona sigui la mare del node. En el cas que aquesta probabilitat superi el 65%, i en el cas que el node no se li hagi assignat prèviament uns altres pares, s'afegeix una relació entre el node i els dos possibles progenitors.

La figura 6 ens mostra el graf de connexions resultant de la base de dades familiar, amb les connexions ja fetes.

5.2 Creació de l'arbre

Un cop s'ha trobat tot el graf de relacions, amb les connexions incloses, s'ha procedit a implementar l'algorisme de cerca per tal de buscar l'arbre genealògic.

Per buscar els pares d'un node s'ha decidit implementar l'algorisme de cerca en amplada o BFS (Breadth First Search). Aquest algorisme comença a partir d'un node arrel, que és el node del qual volem saber l'arbre, i s'explora els seus veïns, en aquest cas els seus pares. A continuació s'exploren els veïns dels pares i així successivament fins que els nodes no troben cap veí, és a dir, que no tenen pares.

Per a la realització d'aquest arbre s'ha creat un nou graf on s'introdueixen tots els nodes que l'algorisme va analitzant, amb les seves respectives connexions. La figura 7 ens permet observar el resultat de l'arbre que ens retorna el programa.

Al mateix moment que es va creant l'arbre, també es crea la llista dels nodes ordenats segons el sistema Sosa-Stradonitz (explicat al punt 6).

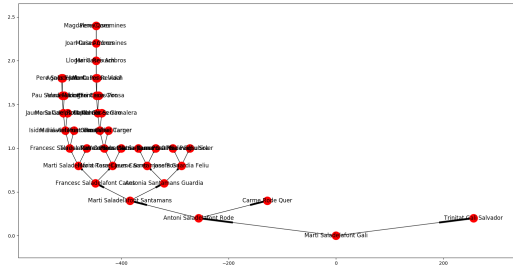


Fig. 7: Arbre genealògic a partir de la base de dades familiar

6 IMPLEMENTACIÓ DE L'APLICACIÓ

Per tal de poder realitzar les cerques de manera molt més amigable i fàcil, s'ha procedit a la implementació d'una aplicació d'escriptori. Per a tal efecte, s'ha utilitzat la llibreria de PythonTk.

Una de les condicions de la realització d'aquesta aplicació és que ha de ser intuïtiva, de fàcil ús. Per tant, s'ha dissenyat de manera que qualsevol usuari pugui buscar la informació de manera ràpida.

La figura 8 ens mostra com és la versió final de l'aplicació.

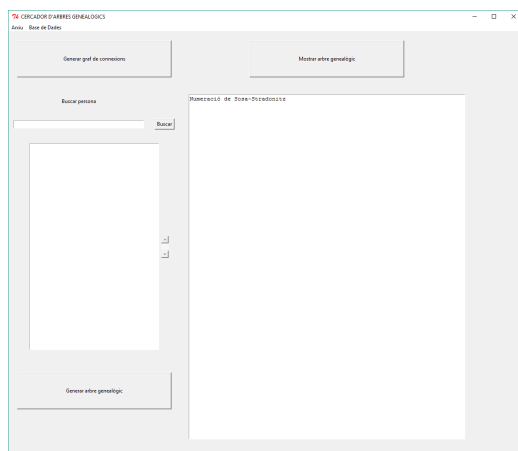


Fig. 8: Aplicació d'escriptori

Com es pot veure, el botó *Generar graf de connexions* ens permet generar el graf amb totes les diferents connexions que tenen els nodes. En el cas que la base de dades que s'està utilitzant ja ha estat utilitzada amb anterioritat, el graf es generarà a partir de la taula *connexions* (explicada a

l'apartat 3.4). En el cas contrari, es generarà el graf a partir dels registres originals de la base de dades.

Un cop s'ha generat el graf, s'ha implementat un buscador de persones. S'ha realitzat per tal de buscar a la base de dades la persona de què volem saber l'arbre genealògic de forma ràpida. L'usuari introdueix el nom o cognom de la persona que busca i el programa mostra un llistat amb les persones que coincideixen amb la cerca. Aquest llistat ens mostra, com podem comprovar a la figura 9, l'identificador personal i el nom de la persona.

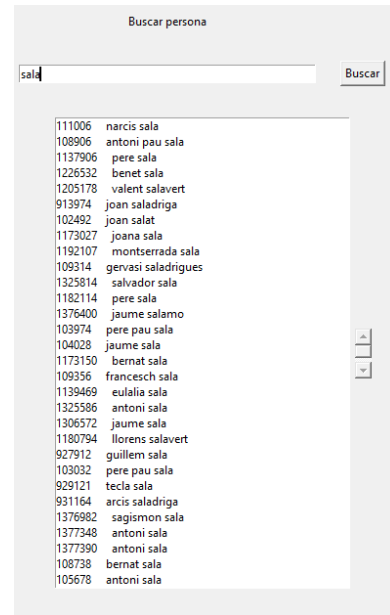


Fig. 9: Buscador de persones

Un cop seleccionada la persona interessada, es procedeix a la generació de l'arbre genealògic. Per això s'ha afegit el botó que ho realitza.

Després de generar l'arbre genealògic de la persona interessada, podem observar l'arbre resultant (tal com ens mostra la figura 7). Per aquest fet s'ha afegit el botó *Mostrar arbre genealògic*. Per completar la informació del node, s'ha afegit una llista dels avantpassats seguint la numeració Sosa-Stradonitz.

Sosa-Stradonitz és un sistema de numeració d'avantpassats en una genealogia ascendenta. Aquest sistema consisteix a assignar el número 1 a la persona del qual s'exposa la genealogia, el número 2 al seu pare i el número 3 a la seva mare. A cada home se li assigna el doble del número del seu fill o filla ($2n$) mentre que a cada dona se li assigna el doble més 1 ($2n + 1$). Així, segons aquest sistema de numeració, cada home tindrà nombres parells mentre que cada dona en tindrà d'imparells.

Per a cada node, la llista ens mostra el seu nom, el lloc de naixement, el lloc de residència, la professió, la data de matrimoni i el lloc on s'ha realitzat i el nom dels seus pares. La figura 10 ens mostra com finalment queda la llista Sosa-Stradonitz.

Numeració de Sosa-Stradonitz	
1	<p>JOAN CASAS REXACH Nascut a St Vicens de Calders Habitant a St Vicens de Calders Treballa de Pare: Llogari Casas Ambros Mare: Maria Rexach</p>
2	<p>LLOGARI CASAS AMBROS Nascut a St Vicens de Calders Habitant a St Vicens de Calders Casat el 1685-05-15 a Sant Vicenc de Calders, bisbat de Vic amb Maria Rexach Treballa de Pages Pare: Joan Cases Mare: Maria Ambros</p>
3	<p>MARIA REXACH Nascuda a Habitant a Casada el 1685-05-15 a Sant Vicenc de Calders, bisbat de Vic amb Llogari Casas Ambros</p>
4	<p>JOAN CASES Nascut a St Vicens de Calders Habitant a St Vicens de Calders Casat el 1662-11-28 a Sant Vicenc de Calders, bisbat de Vic amb Maria Ambros Treballa de Pages</p>
5	<p>MARIA AMBROS Nascuda a St Vicens de Calders Habitant a St Vicens de Calders Casada el 1662-11-28 a Sant Vicenc de Calders, bisbat de Vic amb Joan Cases</p>

Fig. 10: Llista Sosa-Stradonitz

S'ha habilitat un botó que genera un arxiu PDF, de nom la persona de la qual hem generat l'arbre genealògic, on es guarda tota la informació que mostra la llista Sosa-Stradonitz.

Per últim, s'ha realitzat dos menús d'opcions. El primer menú consisteix en els diferents pesos que es donen als paràmetres per poder decidir les diferents connexions del graf (per seleccionar el pare i la mare del node). S'ha habilitat per tal de poder-los modificar en cas que es vulgui donar més pes a un paràmetre que a un altre. S'ha de tenir en compte que la suma dels pesos que s'utilitzen per als dos càlculs (els pesos per calcular els possibles pares i els pesos per calcular la mare) han de sumar 1, ja que el que es calcula és una probabilitat.

El segon menú d'opcions fa referència a la base de dades que s'està utilitzant. En ell es pot configurar els paràmetres (nom d'usuari, contrasenya) de l'entorn MySQL i la base de dades de la qual volem generar el graf.

7 RESULTATS OBTINGUTS

En aquesta secció s'explicarà els resultats obtinguts, tant pel que fa als grafs que es generen com als resultats obtinguts en les diferents bases de dades.

7.1 Grafs

Els grafs que s'havien de generar per a la realització correcta d'aquest treball s'han generat de forma satisfactòria.

El principal problema que s'ha trobat ha estat la gran quantitat de dades que s'havien d'analitzar per tal de generar aquests grafs. Això ha fet que a l'hora de generar els grafs, sobretot en el cas del graf de connexions, es trigui força temps.

També s'ha trobat algun problema a l'hora de calcular les connexions. Per a la realització del còmput, s'utilitza com a paràmetre les poblacions d'origen i de residència dels nodes. Per a facilitar el càlcul de similitud entre aquestes poblacions, s'ha utilitzat la versió normalitzada del text. A la versió de prova de la base de dades d'Esposalles, que és la que s'ha utilitzat per tal de comprovar correctament el funcionament del programa, aquests dos camps (població d'origen i de residència normalitzades) no estaven emplenats. Així doncs s'ha hagut d'utilitzar les versions no normalitzades, fet que dificultava la correcta realització de la connexió.

Malgrat aquests problemes, el funcionament del programa en relació a la creació dels grafs ha estat correcte.

7.2 Bases de dades

Per analitzar els resultats obtinguts respecte a les bases de dades utilitzades s'ha de tenir en compte que, al principi de la realització dels registres de matrimoni a les diferents parròquies, la informació que es guardava a cada registre de matrimoni era força escassa. Normalment s'hi posava la informació bàsica del marit (nom i cognom) així com la població d'origen i la professió, tot i que aquest últim no s'hi incloïa sempre. Pel que fa a la muller, no sempre s'hi indicava el seu cognom, ja que normalment se solia incloure només el nom.

Partint d'aquesta premissa, els resultats obtinguts utilitzant la base de dades d'Esposalles que s'ha facilitat són els esperats teòricament. Com que aquesta base de dades correspon als matrimonis compresos entre 1575 i 1619, les dades en relació a algun dels matrimonis i de les persones que en formen part són escasses. Així doncs, es podria esperar que el programa trobés dificultats a l'hora de trobar una connexió entre els diferents nodes. I això és el que s'ha trobat a l'hora de comprovar els resultats.

Per altra banda, el fet que es disposi únicament de la informació relacionada amb el matrimoni també dificulta la certesa que la connexió és real, és a dir, que el pare és realment el pare del node. Per això aquesta connexió es calcula d'acord amb probabilitats.

Malgrat això, la majoria de les connexions que s'han creat utilitzant la base de dades d'Esposalles podríem considerar que són connexions correctes, amb una alta probabilitat, ja que en la majoria de casos tant el cognom del pare com el del fill coincideixen, així com les poblacions d'origen i destí i en algun cas fins i tot també l'ocupació.

Si ens fixem amb la base de dades familiar, al tenir molta més informació de cada un dels contraents del matrimoni (com ara segons cognoms de cada contraent) fa que les diferents connexions es creïn amb molta més exactitud.

També ha estat molt útil la utilització d'aquesta base de dades per a la comprovació del correcte funcionament del programa, ja que al tenir ja recollit l'arbre genealògic correcte la comprovació amb el resultat obtingut de l'execució del programa ha estat més senzilla.

8 CONCLUSIONS

Un cop finalitzat aquest treball, podem extreure com a conclusions que s'ha assolit correctament els principals objectius marcats per a la realització d'aquest TFG.

S'ha seguit correctament la planificació establerta a l'inici del treball, seguint correctament els períodes plantejats per a cada tasca o subtasca. Seguir la metodologia que s'ha comentat ha ajudat sobre manera a poder seguir la planificació al detall. El fet de realitzar reunions periòdiques amb el tutor, per a resoldre dubtes sobre l'enfocament del TFG també hi ha ajudat.

La realització de l'aplicació ha estat correcta, fent que aquesta sigui de fàcil ús per a qualsevol usuari, clara i entenedora. De cara al futur, seria un complement ideal per a la correcta utilització d'aquesta eina la realització d'un manual d'usuari.

Malgrat que s'havia comentat amb el tutor, no s'ha pogut realitzar la implementació de l'ús de bases de dades dels padrons municipals en aquesta eina. De cara a una actualització futura de l'aplicació, seria interessant incloure també el fet de poder utilitzar aquestes bases de dades.

També, com a visió de futur de l'aplicació, seria interessant adaptar-la també a la utilització de registres de naixement i defunció, per a poder complementar la informació dels matrimonis. Això facilitaria la realització de les diferents connexions, fent que siguin més exactes del que l'aplicació actual permet.

AGRAÏMENTS

Aquest projecte s'ha dut a terme amb la supervisió de Josep Lladós Canet. M'agradaria expressar el meu agraïment, ja que la seva ajuda a l'hora d'enfocar correctament el treball i al llarg de la realització d'aquest ha estat molt gran.

També agrair al Joan Mas la seva ajuda en relació amb la base de dades d'Esposalles, pel seu temps de dedicació a resoldre els petits dubtes que he tingut.

REFERÈNCIES

- [4] Buidatges de matrimonis. <http://www.scgenealogia.org/projectes/enxarxeu-ho/buidatges.html>.
- [5] Centre de Visió per Computador. <http://www.cvc.uab.es/>.
- [6] FamilySearch.org. <https://familysearch.org/>.
- [7] Geneanet.com. <http://www.geneanet.org/>.
- [8] MyHeritage.org. <http://www.myheritage.es/>.
- [9] MySQL.com. <https://www.mysql.com/>.
- [10] NetworkX.github.io. <https://networkx.github.io/>.
- [11] Gramps Project. <https://gramps-project.org/introduction-WP/>.
- [12] Python.com. <https://www.python.org/>.
- [1] Ancestry.com. <http://www.ancestry.com/>.
- [2] Sistema General de Documentació familiar. <http://www.gdsystem.net/webv2/ca/>.
- [3] Societat Catalana de Genealogia. <https://www.scgenealogia.org/>.

