Títol: Eines per a l'anàlisi del Càncer a Catalunya

Volum: I

Alumne: Núria Trench Ribes

Directors: Josep Lluís Larriba Pey

Josepa Ribes Puig

Departament: Arquitectura de Computadors (AC)

Data: 13/10/08

SECRETARIA ACADÈMICA FIB
ENTRADA
- 3 NOV. 2008

Facultat d'Informàtica de Barcelona
Universitat Politècnica de Catalunya

PROJECTE FINAL DE CARRERA

Eines per a l'anàlisi del Càncer a Catalunya

Alumna:
Núria Trench Ribes

Directors:
Josep Lluis Larriba Pey
Josepa Ribes Puig

Agraïments

Primer de tot, m'agradaria expressar el meu sincer agraïment a totes aquelles persones que han fet possible la realització d'aquest projecte.

Al Josep Lluís Larriba Pey, el meu tutor, per oferir-me la possibilitat de dur-lo a terme i fer-ne el seguiment amb tan suport.

Al Norbert Martínez i al Sergio Gómez, companys de DAMA-UPC per la seva constant supervisió i interès que hi han dipositat.

A la Josepa Ribes Puig, la meva tia i la codirectora d'aquest projecte, per la seva impagable dedicació, tan a nivell professional com a nivell personal, procurant que en cap moment em manqués res.

A tots els meus companys de DAMA-UPC, l'Aleix, l'Alejandro, l'Arnau, el Cesc, el Joan, el Miquel, el Pere, el Robert, els que han arribat nous i els que ja no hi són, per fer més amenes les hores en aquella sala.

A l'equip del Registre Hospitalari de Tumors de l'Hospital Universitari de Bellvitge i de l'Institut Català d'Oncologia, per l'aportació de valuosa informació i pel seu acolliment.

Als meus amics, per fer més entretinguts aquests mesos de treball.

Al Marcel i la Teresa, els meus pares, pel seu amor i l'incansable suport moral i econòmic sense el qual no hauria estat possible realitzar aquesta carrera i, molt menys, acabar-la; de la mateixa manera que vull agrair-los-hi a la resta de la meva família, especialment, a la meva germana Anna i els meu avis Josep i Pepita, per la seva proximitat malgrat la distància.

I finalment a l'Andreu, el meu xicot, pel seu desbordant optimisme que intenta encomanar-me i, sobretot, per estimar-me de tan a prop.

Índex

1 Idea del projecte **11**

1.1 Motivació 11

1.2 Definició 12

1.3 Situació inicial 13

2 Context del projecte **15**

2.1 Introducció a l'oncologia 15

2.2 Registre de Càncer de Catalunya 16

2.3 Bases de Dades de l'Hospital Universitari de Bellvitge i de l'Institut Català d'Oncologia . . 16

2.3.1 Registre Hospitalari de Tumors 16

2.3.2 Altes Hospitalàries 17

3 Visió general del projecte **19**

3.1 Descripció 19

3.1.1 Introducció 19

3.1.2 Funcionament de l'aplicació 21

3.2 Definició de les consultes 24

3.2.1 Introducció 24

3.2.2 Consulta 1 25

3.2.3 Consulta 2 27

3.2.4 Consulta 3 28

3.2.5 Consulta 4 30

3.2.6 Consulta 5 32

3.3 Flux de dades i informació 34

3.3.1 Base de dades relacional 35

3.3.2 Base de dades orientada a graf 36

3.4 Principal software a utilitzar 38

3.4.1 Llibreria *Jdex* 38

3.4.2 L'eina de visualització *Prefuse* 40

4 Anàlisi i Disseny **41**

4.1 Anàlisi 41

4.1.1 Anàlisi de requisits 42

4.2 Especificació 46

4.2.1 Model de casos d'ús 47

4.2.2 Model conceptual de la Base de Dades Relacional . . . 55
4.2.3 Model conceptual del sistema software . . . 57
4.3 Disseny . . . 61
4.3.1 Arquitectura . . . 61
4.3.2 Patró arquitectònic . . . 61
4.3.3 Model de comportament. Diagrames de seqüència . . . 62

5 Implementació **69**
5.1 Introducció . . . 69
5.1.1 Llenguatge de programació . . . 69
5.1.2 Llibreries . . . 70
5.2 Capa de dades . . . 70
5.2.1 Introducció . . . 70
5.2.2 Base de Dades relacional . . . 70
5.2.3 Base de dades orientada a graf . . . 77
5.2.4 Càrrega de la Base de Dades orientada a graf . . . 77
5.2.5 Fitxers de configuració . . . 78
5.3 Capa de domini . . . 79
5.3.1 Navegació per la Base de Dades orientada a graf . . . 79
5.3.2 Construcció del graf resultant . . . 80
5.3.3 Cost de la construcció del graf resultant . . . 82
5.3.4 Definició de l'exportació del graf resultant . . . 83
5.4 Capa de presentació . . . 83
5.4.1 Exportació del graf resultant . . . 83
5.4.2 Visualització del graf exportat . . . 85
5.4.3 Fugues de memòria . . . 86

6 Planificació i costos **87**
6.1 Planificació del projecte . . . 87
6.1.1 Diagrama de Gantt . . . 88
6.2 Costos del projecte . . . 89
6.2.1 Costos de personal . . . 89
6.2.2 Costos de material . . . 90
6.2.3 Cost total . . . 91

7 Conclusions **93**
7.1 Implicació personal . . . 93
7.2 Coneixements previs . . . 94
7.3 Coneixements adquirits . . . 95

8 Resultats i Futur del Software **97**
8.1 Període de proves . . . 97
8.2 Anàlisi de resultats . . . 97
8.3 Futur del software . . . 98
8.3.1 Continuïtat del projecte . . . 98

8.3.2 Possibles ampliacions 98
8.3.3 Aplicació en altres especialitats 99

A Fitxers de resultats **101**
A.1 Introducció 101
A.2 Fitxer de resultats en format textual 101
A.3 Fitxer de resultats de Microsoft Excel 104
A.4 Fitxer de resultats en format XML 105
A.5 Llista de pacients en format textual 106
A.6 Llista d'ingressos en format textual 107

B Manual d'usuari **109**
B.1 Introducció 109
B.2 Requisits del sistema 109
B.3 Gestió de projectes 109
B.3.1 Crear projecte 110
B.3.2 Obrir projecte 111
B.3.3 Eliminar projecte 111
B.3.4 Desar resultats 112
B.3.5 Tancar projecte 112
B.4 Execució de consultes 112
B.4.1 Execució de la consulta N 113
B.4.2 Visualització del graf resultant 114
B.5 Llegenda de la visualització del graf 119
B.6 Visualització del graf 119
B.6.1 Visualitzar el graf sencer 121
B.6.2 Visualitzar el graf resum 121
B.6.3 Modificar distància amplada 122
B.6.4 Modificar distància profunditat 122
B.6.5 Opcions de filtratge 123
B.7 Manipulació de la visualització del graf 124
B.7.1 Desplaçar àrea visualització 124
B.7.2 Modificar àrea visualització 125
B.7.3 Encabir el graf 125
B.7.4 Seleccionar l'àrea de visualització 126
B.7.5 Expandir node 126
B.7.6 Mostrar informació node 126
B.8 Veure altres resultats 127
B.8.1 Format textual 127
B.8.2 Format full de càlcul 128
B.8.3 Cercar persones 128
B.8.4 Cercar ingressos 128
B.9 Cerca de nodes i visualització de l'etiqueta de nodes i arestes 129
B.9.1 Visualització etiqueta node i aresta 129

B.9.2 Cerca de nodes 129
B.10 Panells informatius 130
B.10.1 Panell informatiu de tasques 130
B.10.2 Panell informatiu de consultes 130
B.10.3 Panell informatiu de projecte 130
B.11 Sortir del programa 131

C Terminologia **133**

D Acrònims **135**

Capítol 1

Idea del projecte

Aquest primer capítol permet situar els actors que intervindran en el projecte i quins són els objectius que cal haver complert un cop hagi estat finalitzat. Principalment doncs, es dóna una petita definició que ajudarà a introduir-nos i situar-nos en la temàtica del projecte i el seu futur desenvolupament.

1.1 Motivació

Els estudis d'observació i predicció de casos diagnosticats de càncer a Catalunya [11] justifiquen la necessitat de reforçar la qualitat assistencial en l'atenció sanitària dels pacients implicats. La planificació sanitària és el procés que té com a objectiu millorar l'assistència sanitària dels pacients. De fet, tota planificació parteix de sistemes d'informació que contenen les dades necessàries per tal d'analitzar els processos sanitaris que s'han dut a terme, determinar les necessitats assistencials dels pacients i definir la utilització dels serveis sanitaris [10]. A més, aquest conjunt de dades, permet avaluar l'èxit de la planificació objectivament mitjançant l'ús d'indicadors preestablerts.

La idea del projecte sorgeix a partir de la necessitat, per part d'un grup que pertany al Pla Director d'Oncologia (PDO) de la Direcció General de Planificació i Avaluació del Departament de Salut, de dur a terme l'anàlisi i avaluació dels serveis sanitaris utilitzats en els procediments diagnòstics i terapèutics registrats a la Base de Dades de les Altes Hospitalàries (AH). Les AH contenen les dades que fan referència a tots els ingressos hospitalaris corresponents als pacients que han presentat, almenys, un diagnòstic de càncer en un centre sanitari determinat. Entre la gran quantitat de dades que hi ha registrades per a cada ingrés hi podem trobar les que determinen el motiu de l'ingrés, els procediments diagnòstics i terapèutics aplicats i les complicacions i/o efectes adversos que han aparegut durant l'ingrés.

Cal dir que, el grup que desenvolupa el rol de client, antigament havia format part de l'Institut Català d'Oncologia (ICO) i, des d'aleshores, s'encarrega del Registre Hospitalari de Tumors (RHT) de l'Hospital Universitari de Bellvitge i de l'Institut Català d'Oncologia [25]. El RHT identifica els casos de càncer assistits en els esmentats centres sanitaris, de manera automàtica, gràcies a les dades contingudes a les AH i a la Base de Dades d'Anatomia Patològica (APA) [12]. El procés automàtic d'identificació dels casos de càncer es realitza mitjançant un software anomenat ASEDAT [3], que recull la data de diagnòstic, la localització tumoral, la histologia i el mètode per assolir el diagnòstic del tumor. Aquest conjunt de dades s'obtenen a partir d'unes regles i algorismes reconegudes internacionalment [2] i són la informació bàsica recomanada per a un RHT.

Per tant, queda clar que ASEDAT és una eina que permet extreure les dades adients per a realitzar l'avaluació dels serveis sanitaris implicats en els ingressos de malalts oncològics i, com a conseqüència,

facilita la planificació sanitària. Ara bé, ASEDAT no disposa de regles i algorismes que permetin recollir, automàticament, les dades de les AH referents als procediments diagnòstics i terapèutics aplicats a cadascun dels pacients malalts de càncer durant els seus ingressos.

Així doncs, l'interès d'avaluar els serveis sanitaris utilitzats en els procediments registrats en les altes hospitalàries dels malalts de càncer i la mancança d'una eina que permeti obtenir les dades necessàries per a dur a terme aquesta avaluació són la motivació principal del projecte.

Per tant, l'objectiu és oferir un sistema software que construeixi informació útil a partir de les AH per tal d'analitzar els procediments diagnòstics i terapèutics aplicats en els ingressos corresponents a cada tipus de tumor diferent. A partir d'aquesta informació, l'usuari ha de poder identificar i definir regles i algorismes que, posteriorment, seran introduïdes a l'ASEDAT amb la finalitat d'automatitzar la tasca de recollida de dades referents a procediments.

La importància de la realització d'aquesta nova eina rau en el fet que, actualment, si es realitzés aquesta tasca de manera manual, implicaria la lectura de tots els ingressos de cada pacient d'ençà que se'ls hi va diagnosticar un càncer. En aquest cas, després de dedicar força temps a llegir un reduït nombre d'ingressos, els investigadors podrien adquirir cert coneixement pel què fa el comportament de cada tumor segons els procediments registrats en cada ingrés i, per tant, seria possible dur a terme una identificació més o menys acurada de les regles i algorismes per a ser introduïts a l'ASEDAT.

És evident que aquest mètode de treball no és ni eficient ni eficaç atesa la gran dedicació i implicació de temps que comporta. A més, el coneixement adquirit d'aquesta anàlisi manual seria aproximat i, conseqüentment, caldria disposar d'un període de temps excessiu per tal d'arribar a algun punt interessant.

És per aquest motiu que l'eina a implementar ha de construir i visualitzar la informació adequada per tal d'avaluar els procediments aplicats en totes les altes hospitalàries relacionades amb un tipus de tumor i, per tant, permetre'n la construcció manual de les regles i algorismes a introduir a l'ASEDAT.

Finalment, el grup del DPO decideix dirigir-se a DAMA-UPC [17] amb el propòsit de presentar-los-hi el problema i cercar, entre les dues parts, una solució tecnològica.

1.2 Definició

L'objectiu principal d'aquest projecte és la realització d'una eina informàtica que permeti construir i ordenar informació útil referent als diferents procediments diagnòstics i terapèutics aplicats als pacients amb càncer de Catalunya. És a dir, permet facilitar l'avaluació objectiva d'aquests procediments continguts a les AH corresponents als malalts oncològics d'un o més centres sanitaris, amb la finalitat de definir regles i algorismes per a ser introduïts a l'ASEDAT [3].

Per a construir la informació esmentada és necessària la manipulació de les dades contingudes en sistemes d'informació ja existents a Catalunya. Les dades que fan referència als ingressos hospitalaris estan degudament estructurades, codificades i informatitzades tal com ho indica l'especificació del registre del Conjunt Mínim Bàsic de Dades (CMBD) [1].

L'objectiu principal es descompon en dos de més específics. Per una banda, l'aplicació final ha de permetre l'obtenció d'informació útil referent als procediments registrats durant els ingressos corresponents a malalts de càncer amb la finalitat de facilitar l'avaluació d'aquests resultats i extreure regles i algorismes per a l'ASEDAT i, per l'altra, ha de disposar de la visualització d'aquesta informació en una estructura en forma de graf i/o en altres formats que representin els resultats d'una manera clara, ordenada i intuïtiva que en garanteixin l'èxit.

1.3 Situació inicial

El projecte neix de l'acord mutu entre el grup del PDO i DAMA-UPC [17] que accepta l'encàrrec de construir un nou software per tal de solucionar la problemàtica de la falta de regles i algorismes a introduir a l'ASEDAT [3] amb la finalitat de recollir, automàticament, les dades referents als procediments aplicats a cada ingrés registrat a les AH.

Entre les diferents tasques que duen a terme els investigadors que formen part del grup del PDO que desenvolupa el rol de client en aquest projecte, s'hi troben la col·laboració amb xarxes nacionals i internacionals de Registres del Càncer, la gestió del RHT [25] de l'Hospital Universitari de Bellvitge i de l'ICO i la construcció del Registre de Càncer de Catalunya.

D'altra banda, el grup de DAMA-UPC està format per un conjunt de desenvolupadors i investigadors que treballen en la recerca i subministrament de tecnologia relacionada amb la gestió i accés a grans volums de dades, la qualitat de la informació i l'exploració d'aquesta. Al mateix temps, realitza projectes en col·laboració amb diferents institucions nacionals i internacionals amb l'objectiu de dur a terme treballs de recerca i desenvolupament. Actualment, el grup del PDO, que anteriorment formava part de l'ICO, és una de les diferents institucions amb qui DAMA-UPC manté projectes de col·laboració des de fa uns anys.

La situació inicial del projecte es centra en el desenvolupament d'una aplicació sense partir de cap implementació prèvia. Així doncs, cal dur a terme un procés d'anàlisi, especificació i disseny complerts per tal de definir detalladament el procés de construcció de l'eina requerida pel client.

D'aquesta manera, es parteix del coneixement adquirit pels professionals que treballen al grup del PDO, ja que han dut a terme l'estudi manual de les dades durant anys. La informació que poden subministrar per tal de dissenyar la nova eina és fonamental per la definició dels objectius, funcionalitats i requisits. És evident que, sense el treball previ d'aquest grup, no seria possible definir la visió global del sistema software que s'ha realitzat en aquest projecte.

També es disposa de les dades que conformen el RHT, recollides gràcies a l'ASEDAT, i les AH. Ambdues, formen part de l'Hospital Universitari de Bellvitge i de l'ICO, i seran les dades de les quals es partirà durant el desenvolupament de la nova aplicació a implementar.

Així doncs, la informació que caldrà mostrar en l'aplicació final vindrà determinada per les exigències del client.

Capítol 2

Context del projecte

Gràcies al primer capítol ha estat possible obtenir una idea general del projecte i donar a conèixer els elements que hi intervenen. No obstant, el fet que aquest projecte es desenvolupi a petició d'un client que pertany a una institució sanitària fa que tots aquells aspectes teòrics no relacionats amb el món de l'enginyeria informàtica siguin de difícil comprensió. Així doncs, per tal d'entendre millor la motivació del projecte i els resultats que es generaran mitjançant la futura aplicació, és interessant aprofundir en l'àmbit de l'oncologia i tots aquells temes relacionats amb la idea del projecte.

2.1 Introducció a l'oncologia

L'oncologia és la part de la patologia que tracta els tumors benignes i malignes, posant especial atenció en els malignes. Els tumors malignes reben el nom de càncer o neoplàsia maligna.

Les diferents tasques que es duen a terme en l'àmbit de l'oncologia són:

- El diagnòstic de tumors benignes i malignes.
- El tractament del càncer mitjançant la cirurgia i teràpies no quirúrgiques, com són la quimioteràpia, radioteràpia i altres modalitats.
- El seguiment de pacients de càncer després d'indicar-los-hi un tractament.
- Les cures pal·liatives dels pacients amb malalties terminals.
- Aspectes ètics relacionats amb l'atenció als malalts de càncer.
- Estudis genètics a la població en general
- Estudis genètics als familiars dels pacients diagnosticats de càncer.
- Estudis epidemiològics amb la finalitat de descriure el càncer a nivell poblacional, identificar poblacions de risc i factors ambientals associats al desenvolupament del càncer.
- Estudis descriptius a la població general sobre la freqüència d'aparició de càncers i els tractaments utilitzats.
- Estudis descriptius a la població general sobre la utilització dels serveis sanitaris per tal de facilitar la planificació sanitària dels departaments de salut d'un territori.

Així doncs, podem veure que l'oncologia és una part de la medicina que treballa tan en investigació com en el tracte directe amb el pacient.

El client d'aquest projecte és un grup format per epidemiòlegs i altres professionals relacionats. De les diferents tasques llistades anteriorment, concretament, se n'encarrega dels estudis epidemiològics i descriptius enumerats en els tres últims punts.

2.2 Registre de Càncer de Catalunya

Actualment, el grup del PDO està involucrat en la creació d'un Registre de Càncer de Catalunya que recopilarà les dades de tots els pacients amb càncer de Catalunya. Les principals dades que formaran part d'aquest registre seran el conjunt de les AH i els serveis registrats a l'APA de tots els centres sanitaris del territori català.

Tot i que aquest registre encara no està construït, s'hi estan dipositant molts esforços per tal de fer-ho possible i cal tenir-lo present com a possible futura Base de Dades de la qual partir en la nova aplicació a desenvolupar. No obstant, en aquest projecte només s'utilitzaran les dades que pertanyen únicament al RHT de l'Hospital Universitari de Bellvitge i de l'ICO, per tal d'oferir la possibilitat de realitzar una primera avaluació dels procediments registrats en aquests dos centres sanitaris.

Així doncs, el Registre de Càncer de Catalunya pretén ser el registre que agrupa les dades oncològiques contingudes en tots els centres de Catalunya, la qual cosa permetrà obtenir una Base de Dades molt potent per a la investigació i per a la planificació sanitària. El software que es desenvolupa en aquest projecte té com a objectiu, a llarg termini, poder treballar partint de les dades registrades en tots els centres sanitaris del territori català. Abans però, és necessari validar els resultats mitjançant les dades dels centres sanitaris esmentats.

2.3 Bases de Dades de l'Hospital Universitari de Bellvitge i de l'Institut Català d'Oncologia

Degut al fet que el projecte es realitza a partir de les dades emmagatzemades al RHT i les AH de l'Hospital Universitari de Bellvitge i l'ICO, cal descriure detalladament alguns aspectes relacionats amb aquestes dues Bases de Dades.

2.3.1 Registre Hospitalari de Tumors

El RHT [25] és un registre compartit entre l'Institut Català d'Oncologia (ICO) i l'Hospital Universitari de Bellvitge (HUB), gestionat pel grup del PDO que desenvolupa el rol de client en aquest projecte. L'objectiu d'aquest registre és emmagatzemar tots els diagnòstics de càncer que s'esdevenen en els centres sanitaris esmentats. ASEDAT [3] és l'encarregat de construir-lo de manera automàtica, gràcies a un conjunt de regles i algorismes que permeten extreure les dades necessàries de les AH i les APA d'aquests centres.

Actualment, conté un volum considerable de casos diagnosticats de càncer. L'1 d'octubre de 1990, el RHT comença l'enregistrament dels diagnòstics de càncer als centres sanitaris dels quals pertany. Després de disset anys, al gener del 2007, el RHT ja comptava amb, aproximadament, 63.000 tumors enregistrats.

Finalment cal dir que, per a cada pacient diagnosticat de càncer registrat al RHT, és possible aconseguir les dades que fan referència a tots els ingressos hospitalaris d'aquest pacient posteriors al citat diagnòstic mitjançant l'encreuament entre aquest registre i les AH de l'Hospital Universitari de Bellvitge i de l'ICO.

2.3.2 Altes Hospitalàries

Totes les altes hospitalàries de tots els centres sanitaris de Catalunya es centralitzen en el registre del CMBD [14]. Aquest registre disposa d'una especificació que detalla, per a cada ingrés a enregistrar, quines variables han de ser emmagatzemades i quin format han de seguir.

Les variables que conformen cada alta hospitalària són les següents:

- Variables que identifiquen el pacient
  - codi d'identificació personal (CIP), que és independent del centre sanitari.
  - data de naixement.
  - sexe.
  - municipi i província de residència.
  - història clínica, que és el número d'identificació d'un malalt en un centre sanitari determinat.
- Variables relacionades amb el procés
  - unitat proveïdora de serveis (UP).
  - tipus d'activitat.
  - règim econòmic.
  - circumstància d'admissió.
  - circumstància d'alta.
  - data d'admissió.
  - data d'alta.
- Variables clíniques
  - diagnòstics codificats mitjançant l'estàndard *International Classification of Diseases, 9th revision*(ICD9) [8].
  - procediments codificats mitjançant l'estàndard ICD9.
- Variables específiques de cada sistema que permeten agrupar les valoracions de pacients en el sistema RUG (Grups d'Utilització de Recursos).

Capítol 3

Visió general del projecte

Aquest capítol té com a objectiu definir el projecte des d'un punt de vista més descriptiu sense entrar en profunditat en aspectes que fan referència a l'àmbit mèdic o informàtic. La idea és proporcionar una visió ràpida que permeti entendre els objectius principals del sistema software que cal implementar. En capítols posteriors, es defineix la mateixa informació de manera més detallada i seguint la metodologia que regeix qualsevol documentació d'un projecte relacionat amb l'enginyeria informàtica.

3.1 Descripció

Cal recordar que l'objectiu principal de la nova eina a implementar és facilitar l'avaluació dels procediments diagnòstics i terapèutics aplicats als malalts de càncer registrats al RHT [25], per tal d'obtenir les regles i algorismes que s'introduiran a l'ASEDAT [3] amb la finalitat d'aconseguir les dades adients que permetran desenvolupar la planificació sanitària del territori català.

A continuació, es defineixen les funcionalitats, els objectius fixats i el resultat que es pretén aconseguir en aquest projecte.

3.1.1 Introducció

Aquesta nova aplicació està pensada per a ser executada en un ordinador local utilitzant com a element d'entrada la Base de Dades relacional emmagatzemada a la mateixa màquina. Aquesta Base de Dades conté el total de pacients que van ser diagnosticats d'un únic tumor juntament amb totes les seves altes hospitalàries posteriors a la data d'aquest diagnòstic, a més de contenir altres dades que hi estan relacionades.

A la Figura 3.1, es mostra un exemple de les dades que fan referència a un pacient registrat a la Base de Dades. En aquest cas, en Joan va ser diagnosticat d'un tumor de colon el 10 de Maig del 2000 i hi ha registrades un total de 3 altes hospitalàries posteriors a l'esmentat diagnòstic. Per cada alta hospitalària, es disposa de les dades que fan referència als diagnòstics i procediments diagnòstics i/o terapèutics aplicats.

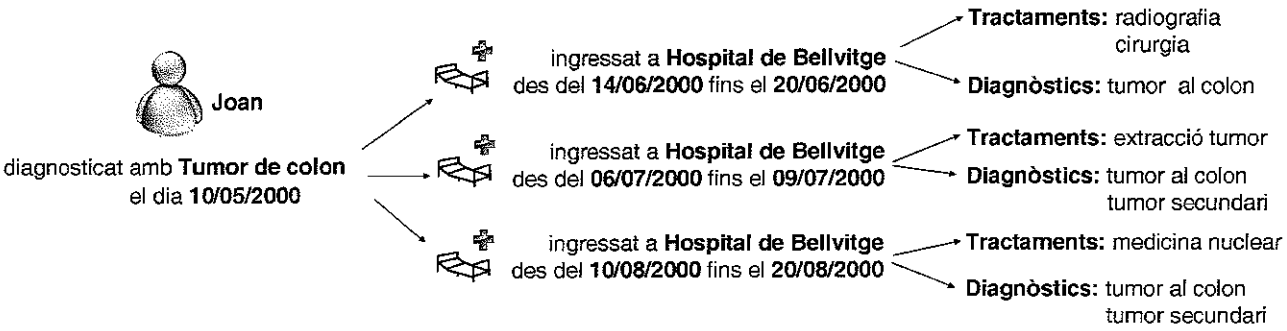


Figura 3.1: Exemple de dades associades a un pacient

A partir d'aquestes dades, l'aplicació pretén oferir un conjunt de consultes contra la Base de Dades amb l'objectiu de construir uns resultats que representin la informació necessària per a avaluar els procediments diagnòstics i terapèutics registrats per a cada ingrés de cada malalt de càncer i, conseqüentment, permetre a l'usuari la identificació i determinació de les regles i algorismes per a l'ASEDAT.

La representació d'aquesta informació és una fase molt important en aquest projecte atesa la gran quantitat de dades a partir de la qual es construeixen els resultats que acostumen a ser de dimensions considerables.

Així doncs, és necessari visualitzar la informació resultant mitjançant recursos que garanteixin una representació clara, entenedora i amigable. Alhora, per a realitzar un estudi còmode dels resultats, cal que aquesta informació sigui fàcilment navegable. És per aquest motiu que s'ha decidit utilitzar una representació en forma de graf que, a més, permeti explotar aquells elements que aportin informació útil i obviar aquells que siguin irrellevants.

Finalment, l'aplicació permet obtenir els resultats en diferents formats per tal de visualitzar la informació de diverses maneres i, així, facilitar l'estudi i donar diferents punts de vista útils per a l'usuari.

El gràfic de la Figura 3.2mostra un resum de les funcionalitats que ofereix l'aplicació i com interaccionen entre elles.

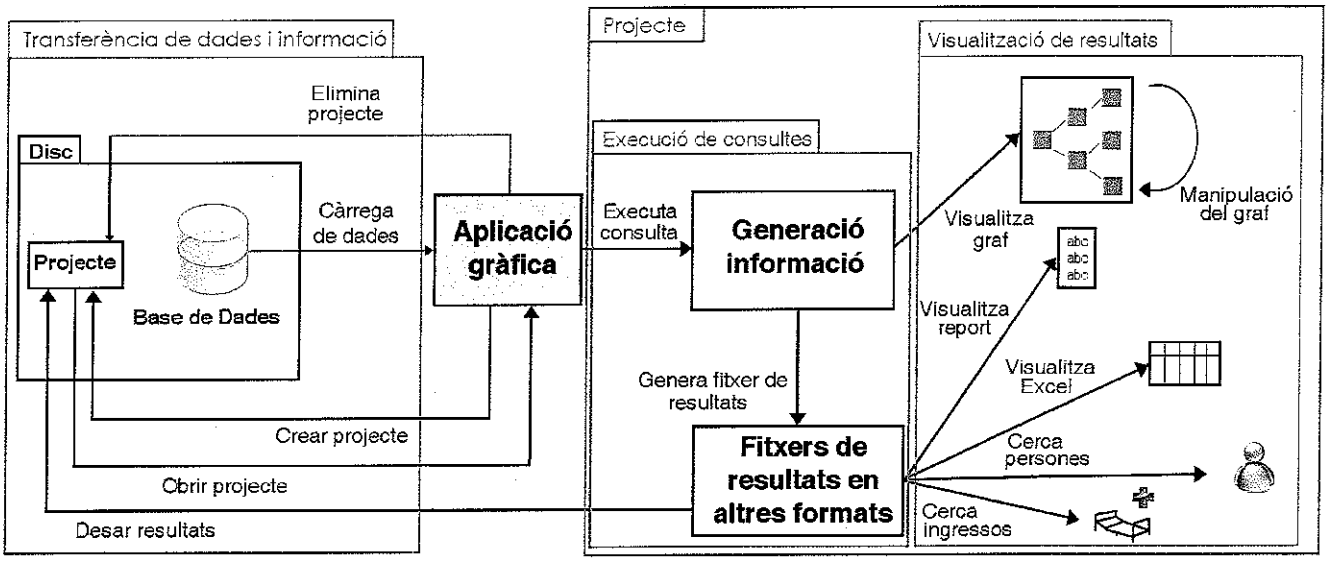


Figura 3.2: Resum de les funcionalitats de l'aplicació

3.1.2 Funcionament de l'aplicació

Després de la petita introducció de les funcionalitats que conformen el sistema software a desenvolupar i que han estat representades en el gràfic de la Figura 3.2, és el moment de descriure quina és la funció de cadascuna d'elles.

3.1.2.1 Introducció

Primer de tot, cal dir que l'aplicació treballa en un entorn de projectes. És a dir, per a poder utilitzar les funcionalitats principals que ofereix, és necessari obrir un projecte creat anteriorment. En el requadre "Projecte" de la Figura 3.2, podem veure quines són aquestes funcions.

Un projecte conté les dades de configuració per a poder executar les diferents consultes i, per tant, per a poder-hi realitzar qualsevol tipus d'operació relacionada. A més, és possible emmagatzemar-hi els fitxers de resultats generats en cada execució de consulta.

Així doncs, les dades que conformen la configuració d'un projecte són les següents:

- Localització tumoral a partir de la qual s'executen les consultes.
- Directori on s'emmagatzema a disc el projecte.
- Directoris on s'emmagatzemen a disc els resultats generats durant les execucions de les consultes.

Per tant, l'anterior llista representa el conjunt de dades que es carreguen en el moment d'obrir un projecte. Gràcies a això, és possible iniciar l'execució de les consultes i realitzar el corresponent anàlisi dels resultats. Cada projecte, tal i com s'ha vist, té un tumor associat a partir del qual es realitzen les consultes per tal d'obtenir informació útil i, poder així, avaluar els procediments registrats per a cada alta hospitalària de cada pacient diagnosticat amb el tumor indicat pel projecte.

Un cop es disposa d'un projecte obert amb les corresponents dades carregades, és possible generar els fitxers amb els resultats de les execucions de les consultes i emmagatzemar-los allà on la configuració del projecte tingui establert.

3.1.2.2 Transferència de dades i informació

Existeixen vàries funcionalitats que permeten la interacció entre aplicació i disc amb l'objectiu de carregar o emmagatzemar el conjunt de dades i informació necessàries pel correcte funcionament de l'aplicació. A la Figura 3.2 es pot veure quines són aquestes funcionalitats.

Concretament, les dades que intervenen en la transferència que s'estableix entre el sistema software i el sistema d'emmagatzemament físic són les següents:

- Conjunt de configuracions, dades i fitxers de resultats emmagatzemats en cada projecte.
- Base de Dades que conté el conjunt de persones diagnosticades d'un tumor juntament amb les seves altes hospitalàries associades.

A continuació, es descriu en detall quin és l'objectiu de cadascuna d'aquestes.

Gestió de projectes

Atès el fet que el sistema treballa en un entorn de projectes, cal descriure els objectius de les tres funcionalitats que interaccionen entre l'aplicació i el disc.

En primer lloc, l'aplicació ofereix la possibilitat de crear un projecte. Tan sols és necessari associar-li un tipus de tumor a partir del qual s'executaran les consultes que es creguin necessàries i que construiran la informació útil per a realitzar l'estudi dels procediments aplicats als ingressos relacionats amb aquest tumor. Des del moment que es disposa d'almenys un projecte emmagatzemat a disc, l'aplicació ja està preparada per obrir un projecte i carregar-ne les dades que fan referència al tipus de tumor associat.

De la mateixa manera que es pot obrir un projecte, l'aplicació permet tancar-lo. Per tant, es poden obrir i tancar tots els projectes desats a disc.

Finalment, cal dir que també és possible eliminar un projecte. Aquest fet implica esborrar tota aquella informació emmagatzemada a disc que hi està relacionada com és la configuració del projecte i els fitxers de resultats que hi havia desats.

Càrrega de les dades

Des de l'instant que existeix un projecte obert i s'executa una consulta, l'aplicació necessita accedir a la Base de Dades. D'aquesta manera, el sistema software carrega les dades pertinents amb l'objectiu de construir-ne els resultats.

Desar els resultats

En cada projecte, l'execució d'una consulta ofereix la possibilitat de generar fitxers de resultats addicionals. En el cas que l'usuari desitgi emmagatzemar aquests resultats a disc, l'aplicació posa a la seva disposició una funcionalitat que permet desar-los en els directoris especificats a la configuració del projecte obert en aquell moment.

Per a cada fitxer de resultats que s'emmagatzema a disc se li associa la data de creació, la consulta que l'ha generat i el tipus d'informació que conté.

3.1.2.3 Execució de consultes

L'execució de consultes és la funcionalitat més important de l'aplicació ja que l'objectiu principal del projecte era oferir una eina per a facilitar la identificació i determinació de regles i algorismes per a ser introduïts a l'ASEDAT mitjançant l'estudi dels resultats obtinguts en cada consulta que permeten l'avaluació dels procediments registrats en les altes hospitalàries de cada pacient diagnosticat d'un càncer en concret.

El sistema software que cal implementar permet executar fins a cinc consultes diferents que comparteixen el mateix objectiu. El conjunt de resultats que s'obtenen per a cadascuna d'elles permet, als investigadors, analitzar detingudament la informació per tal d'extreure'n les regles i algorismes adients. Totes elles realitzen consultes contra la Base de Dades partint de la localització tumoral associada al projecte obert. La informació generada com a resultat es representa mitjançant la visualització d'un graf fàcilment navegable.

A més, per a cada consulta executada, existeix la possibilitat de seleccionar quins arxius addicionals han de ser generats amb la finalitat de mostrar-hi la mateixa informació que la que hi ha representada a la visualització del graf resultant. Cadascun d'aquests fitxers presenta un format diferent per tal de representar la informació de diferents maneres. A la secció 3.3 a la pàgina 34, es descriu detalladament quins són aquests fitxers.

Atesa la importància d'aquesta funcionalitat, s'ha decidit oferir una definició més àmplia dels objectius i elements que conformen les cinc consultes a la secció 3.2 a la pàgina següent.

3.1.2.4 Visualització dels resultats

És important conèixer com es mostren cadascun dels resultats obtinguts durant l'execució de les consultes. La visualització de la informació resultant és el mitjà que permet estudiar els resultats i obtenir les regles i algorismes pertinents per al software ASEDAT.

Visualització del graf resultant

L'objectiu és mostrar la informació construïda durant l'execució de cada consulta mitjançant una visualització en forma de graf. Aquest format de representació permet, a l'usuari, analitzar els resultats d'una manera més intuïtiva gràcies a l'organització de la informació en un graf. A més, possibilita una fàcil navegabilitat a través dels seus elements. És a dir, per una banda, és possible ocultar aquells elements del graf que es considerin irrellevants i tan sols consultar-los quan sigui necessari. D'aquesta manera, s'aconsegueix millorar la visibilitat dels resultats sense haver-hi d'entrar en detall. D'altra banda, la visualització en forma de graf també permet mostrar tota la informació resultant sense obviar-ne cap element.

Alhora, aquest tipus de representació, facilita la tasca de seguir el camí de relacions entre un element del graf i un altre amb l'objectiu de permetre'n l'observació i estudi de la relació que mantenen tots els elements involucrats en el camí esmentat.

Per acabar, l'aplicació permet realitzar cerques sobre la informació representada en el graf resultant, de manera que assenyala en un color diferent aquells elements que coincideixen amb el text introduït.

Manipulació de la visualització del graf

El sistema software ofereix diverses funcionalitats per tal de fer més amena i flexible la visualització del graf, a més de permetre'n la possibilitat de realitzar-hi canvis.

Per una banda, és possible ampliar o reduir la part del graf que es visualitza, així com desplaçar-ne aquesta àrea de visió amb l'objectiu de mostrar altres parts del graf que poden resultar d'interès per a l'usuari.

Alhora, l'aplicació permet canviar la ubicació dels elements que componen el graf, la distància que existeix entre ells i l'encabiment de tot el graf dins l'àrea de visualització.

Finalment, cal dir que existeixen dos tipus diferents de visualitzacions. Per una banda, és possible mostrar tots els elements del graf oferint el màxim detall i, per l'altra, se'n pot veure un resum amb la possibilitat d'explotar aquells elements del graf que siguin d'interès.

Visualització dels fitxers de resultats addicionals

Permet visualitzar la informació continguda en els fitxers de resultats addicionals que han estat generats durant les execucions de les consultes a petició de l'usuari. La visualització d'aquests resultats es realitza mitjançant una nova finestra.

Es pot obtenir des d'un report de la informació resultant d'una consulta fins a la visualització d'un fitxer de *Microsoft Excel* que conté els mateixos resultats que l'anterior però, en aquest cas, estan organitzats en fulls de càlcul.

Finalment, l'aplicació dóna la possibilitat de localitzar les persones i/o ingressos relacionats amb una informació en concret del graf resultant mitjançant el cercador que ofereix per a aquesta funció.

3.2 Definició de les consultes

S'ha implementat un total de cinc consultes diferents. Per a cadascuna d'elles, l'aplicació genera un conjunt de resultats que conté diferents tipus d'informació depenent de la consulta.

A la Figura 3.2 es pot veure quin és el procés que segueix el sistema software durant l'execució de qualsevol d'aquestes cinc. L'aplicació pren el tipus de tumor associat al projecte i executa la consulta contra la Base de Dades partint d'aquesta dada. A partir d'aquí, es carreguen les dades necessàries i es construeix la informació resultant amb el propòsit de mostrar-la en forma de graf dins l'àrea de visualització.

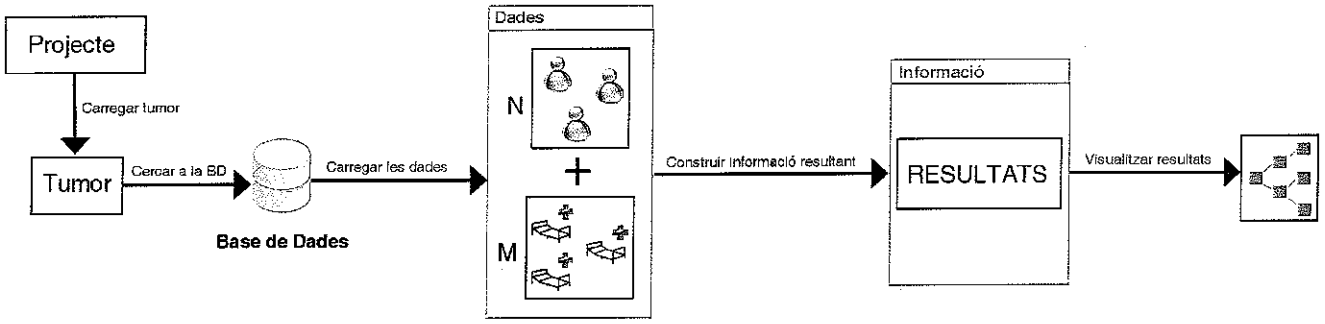


Figura 3.3: Procés durant l'execució d'una consulta

Per a veure una petita definició i un exemple real dels resultats de cadascuna de les cinc consultes, es pot veure la secció 3.2 corresponent al manual d'usuari.

3.2.1 Introducció

És força complicat descriure les dades que precisa cada consulta, la informació que genera i la visualització dels resultats. De fet, és molt més senzill comprendre'n els objectius a través d'un exemple o una demostració mitjançant l'aplicació gràfica. És per aquest motiu que, primer de tot, es definirà aquell vocabulari que pugui generar ambigüitats o malentesos per tal de fer una mica més fàcil la lectura i comprensió d'aquesta secció.

3.2.1.1 Alta hospitalària

És el moment en què, un centre hospitalari, dóna l'alta a un pacient que estava ingressat. En aquest projecte doncs, quan es parla d'una alta hospitalària es fa referència a un ingrés hospitalari en el qual el pacient ha abandonat l'hospital al final de la seva estada.

És possible que una persona que hagi estat diagnosticada d'un tumor no hagi d'ingressar mai després de la data d'aquest diagnòstic. Sovint però, les persones diagnosticades d'un tumor ingressen alguna vegada després de ser diagnosticades. Fins i tot, depenent de les malalties concomitants, les complicacions que puguin sorgir o l'evolució de la pròpia malaltia, un persona pot acumular un bon nombre d'ingressos posteriors a aquest diagnòstic.

3.2.1.2 Procediment

És el conjunt d'accions que es posen en pràctica per tal de diagnosticar, curar o pal·liar la malaltia diagnosticada en un ingrés. Per una banda, existeix el procediment diagnòstic que permet determinar el motiu d'ingrés, i per l'altra, hi ha el procediment terapèutic que permet curar o pal·liar la malaltia diagnosticada durant l'ingrés. Concretament, en cada ingrés, es registra un procediment diagnòstic o terapèutic principal i un conjunt de procediments diagnòstics o terapèutics secundaris.

- **Procediment principal:** Representa aquell procediment que és més important aplicar al pacient ingressat, o bé, el què requereix més urgència. A més, ha d'estar associat al motiu de l'ingrés.

- **Procediments secundaris:** Representen tots aquells procediments que ha estat necessari aplicar al pacient ingressat després d'haver recorregut al procediment principal.

És possible que existeixin casos en els quals no s'hagi registrat cap procediment principal ni secundari. Ara bé, és freqüent que durant un ingrés tan sols s'apliqui un procediment principal sense necessitat d'un secundari. D'altra banda, si no existeix un procediment principal tampoc es registrarà un procediment secundari.

3.2.1.3 Diagnòstic

És la determinació d'una malaltia a partir de la identificació de signes o símptomes que la caracteritzen. De la mateixa manera que s'esdevé pels procediments, en cada ingrés, es registra un diagnòstic principal i un conjunt de diagnòstics secundaris.

- **Diagnòstic principal:** Quan existeix més d'un diagnòstic, és la determinació d'aquell que coincideix amb els màxims signes o símptomes que presenta l'individu, o bé, és el què requereix més atenció atesa la seva urgència. A més, ha de ser el motiu pel qual ha ingressat el pacient.

- **Diagnòstics secundaris:** Són tots aquells diagnòstics que es determinen després d'haver registrat el diagnòstic principal.

En aquesta Base de Dades, és possible que un ingrés no tingui registrat cap diagnòstic. No obstant, si no existeix un diagnòstic principal tampoc n'existirà cap de secundari. Un ingrés pot tenir un diagnòstic principal associat i, en canvi, no disposar de cap diagnòstic secundari.

A continuació, es presenta una definició per a cada consulta i s'exposa quina és la informació representada en cadascuna de les visualitzacions del graf resultant.

3.2.2 Consulta 1

Obté el conjunt de tots els procediments principals i secundaris d'entre totes les altes hospitalàries relacionades amb els individus que se'ls hi va localitzar un tumor determinat en algun moment de la seva vida.

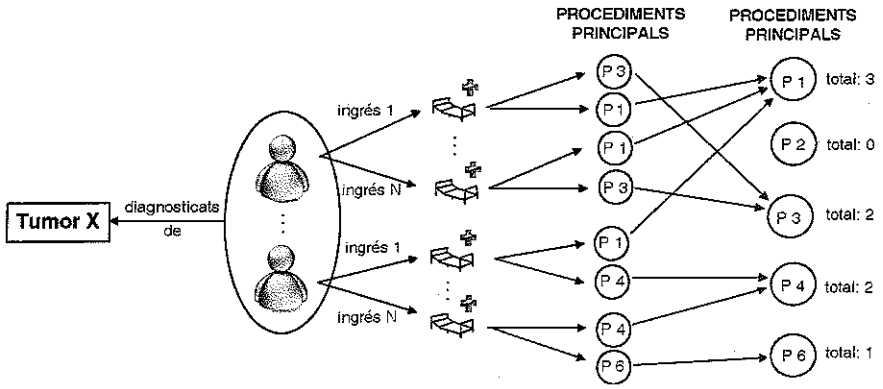


Figura 3.4: Definició de la consulta número 1

Com es pot veure a la Figura 3.4, cal partir del conjunt de persones registrades a la Base de Dades que han estat diagnosticades d'un tumor en concret. Aquest tumor està definit a la configuració del projecte. A partir d'aquí, cal recórrer totes les altes hospitalàries associades a aquestes persones que, a més, han de ser posteriors a la data del diagnòstic. Finalment, per a cadascuna d'aquestes altes hospitalàries, és necessari obtenir el procediment principal juntament amb els procediments secundaris aplicats durant l'ingrés.

L'objectiu de la consulta és determinar els diferents tipus de procediments que s'han aplicat en totes les altes hospitalàries relacionades amb el conjunt de pacients esmentat. Un cop es disposa de cadascun dels diferents tipus de procediments representats de color vermell a la Figura 3.4, és el moment de comptar el total de vegades que han estat aplicats.

3.2.2.1 Utilitat en la investigació oncològica

Els resultats que s'obtenen en aquesta consulta donen una visió general dels diferents tipus de procediments principals i/o secundaris aplicats en els pacients diagnosticats d'un tumor en concret. Gràcies a la visualització del graf, és fàcil fer-se una idea de la quantitat de procediments diferents si es compta el total de nodes que hi ha representats.

D'altra banda, permet veure quin és el procediment que s'aplica més vegades i amb quina freqüència ho fa. A més, és fàcil estudiar aquells nodes que solament agrupen una persona i que, com a conseqüència, representen casos estranys en els quals s'aplica un procediment poc freqüent.

Per tant, és una bona consulta per a iniciar l'avaluació dels procediments diagnòstics i terapèutics aplicats en el conjunt d'ingressos pertanyents als pacients malalts de càncer.

3.2.2.2 Representació dels resultats

La Figura 3.5 mostra com es visualitza la informació que s'obté en l'execució d'aquesta consulta. El node central de color vermell representa el conjunt de les persones que han estat diagnosticades del tumor X. El següent nivell de nodes de color groc, és el conjunt de procediments principals i/o secundaris diferents que s'han extret de totes les altes hospitalàries obtingudes. Aquests nodes, coincideixen amb els procediments dibuixat mitjançant un cercle vermell a la Figura 3.4. Finalment, les arestes que uneixen el node central amb la resta de nodes representen el total de les altes hospitalàries que tenen registrat el procediment representat a l'extrem d'aquesta. Aquesta quantitat, coincideix amb el nombre que apareix a la dreta de cada procediment encercat en color vermell de la Figura 3.4.

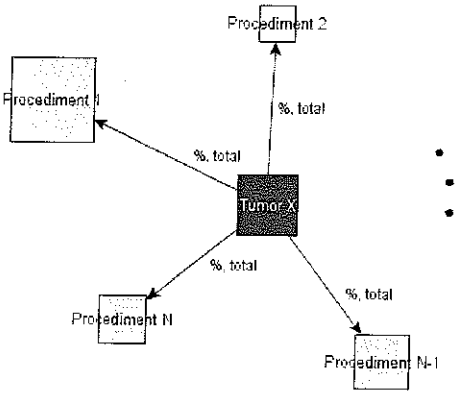


Figura 3.5: Visualització del graf resultant de la consulta 1

3.2.3 Consulta 2

Obté el conjunt de tots els diagnòstics principals i secundaris d'entre totes les altes hospitalàries relacionades amb els individus que se'ls hi va localitzar un tumor determinat en algun moment de la seva vida.

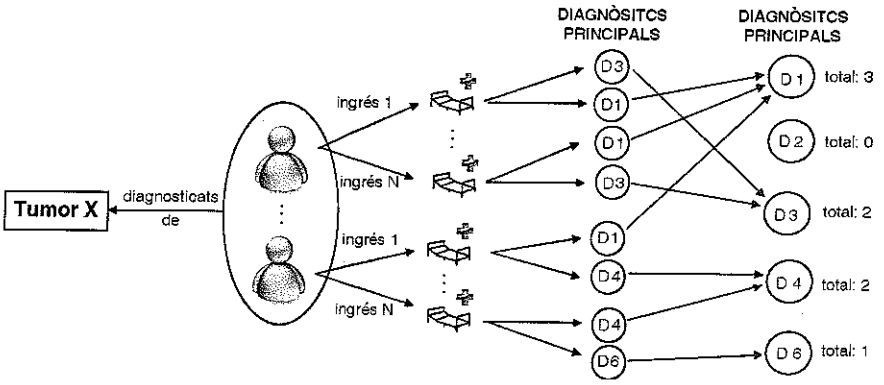


Figura 3.6: Definició de la consulta número 2

Com es pot veure a la Figura 3.6, la definició d'aquesta consulta és molt similar a la què s'ha dut a terme a la primera. La única diferència observable és que, per a cada alta hospitalària que s'obté, ara ens interessa prendre el diagnòstic principal i el conjunt de diagnòstics secundaris.

Consegüentment, l'objectiu varia respecte el de l'anterior. En aquest cas, es tracta de determinar els diferents tipus de diagnòstics registrats que s'han pogut extreure del conjunt d'altes hospitalàries associades als pacients diagnosticats amb el tumor X. A la Figura 3.6, es pot veure clarament com s'agrupen els diagnòstics de les diferents altes hospitalàries dins de cada tipus diferent de diagnòstic representat dins un cercle vermell.

3.2.3.1 Utilitat en la investigació oncològica

Aquesta consulta ha estat dissenyada amb l'objectiu d'oferir una utilitat similar a la primera. Els resultats que s'obtenen permeten tenir una visió general dels diferents diagnòstics registrats en els ingressos d'aquells pacients que van ser diagnosticats d'un tumor en concret. La informació resultant dóna una primera idea dels diagnòstics que es registren durant els ingressos hospitalaris de les persones esmentades i, per tant, també ajuda a la identificació i determinació de regles i algorismes que seran introduïdes a l'aplicació ASEDAT [3].

Després d'haver observat els resultats obtinguts en les primeres dues consultes, ja es disposa de la informació bàsica per avaluar els procediments diagnòstics i terapèutics aplicats en les altes hospitalàries relacionades amb un tipus de tumor. El següent pas és executar les altres tres consultes partint de la mateixa localització tumoral i observar-ne la informació resultant per tal d'acabar de definir i determinar amb més precisió les regles i algorismes esmentades.

3.2.3.2 Representació dels resultats

La descripció de la representació de la informació resultant per a aquesta consulta és pràcticament igual que la utilitzada per a la primera. Així doncs, els nodes de color groc de la Figura 3.7 que abans representaven els diferents tipus de procediments, ara són els diferents tipus de diagnòstics. Aquests nodes equivalen als diagnòstics encerclats en color vermell a la Figura 3.6. Pel que fa la resta d'elements representats, se'ls defineix de la mateixa manera que a la primera consulta.

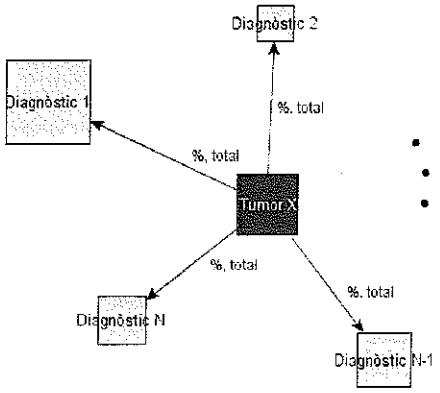


Figura 3.7: Visualització del graf resultant de la consulta 2

3.2.4 Consulta 3

Obté el conjunt de tots els procediments principals d'entre totes les altes hospitalàries relacionades amb els individus que se'ls hi va localitzar un tumor determinat en algun moment de la seva vida, agrupats pel número d'alta hospitalària i ordenats cronològicament.

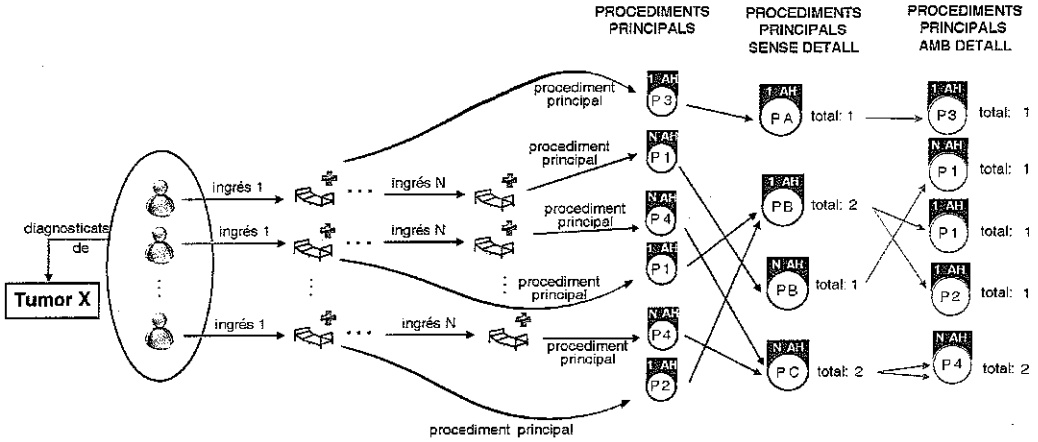


Figura 3.8: Definició de la consulta número 3

És a dir, tal i com s'observa a la Figura 3.8, el què interessa és extreure la relació entre els diferents ingressos hospitalaris de cada pacient. És a dir, cal relacionar els procediments principals aplicats en les primeres altes hospitalàries de cada pacient, i fer el mateix per a totes les altes posteriors de manera que quedin agrupats pel seu número d'ingrés.

Seguint amb l'esquema gràfic de la Figura 3.8, per a cada persona del conjunt que representa el total de pacients diagnosticats d'un tipus de tumor en concret, és necessari obtenir totes les altes hospitalàries associades a cadascun d'ells. A diferència de les dues anteriors consultes, aquesta construeix la informació resultant tenint en compte l'ordre dels diferents ingressos de cada persona. Cal recordar que, en les altres dues, es prenien els diagnòstics o procediments i s'agrupaven sense considerar l'ingrés del qual provenien.

En aquest cas doncs, per a cada alta hospitalària de cada pacient, es pren el procediment principal que se li va aplicar i s'anota a quin ingrés correspon (primer, segon, últim, etc.). Com es pot veure a la Figura 3.8, es poden obtenir procediments des del primer ingrés d'una persona (1 AH) fins a l'últim (N AH). El propòsit és aconseguir agrup en un mateix conjunt tots els procediments principals registrats durant el primer ingrés de cada persona, així com fer el mateix en un altre conjunt per a tots els del segon ingrés fins a acabar agrupant els de l'últim.

Per a disposar d'una millor organització d'aquest conjunt de procediments principals, s'ha optat per agrupar-los en procediments genèrics que engloben procediments de característiques similars. Tal i com es mostra a la Figura 3.8, els procediments principals que s'extreuen del conjunt d'altes hospitalàries representats mitjançant cercles negres, es classifiquen en el grup que representa el procediment genèric al qual pertanyen que, a la Figura 3.8, s'identifica mitjançant un cercle vermell.

Tal i com s'ha dit, l'objectiu de la consulta és determinar els diferents tipus de procediments principals que s'han aplicat de manera que quedin agrupats per primer ingrés, segon ingrés, etc. fins arribar a l'últim, de les persones diagnosticades d'un tumor en concret. Alhora, es presenta la quantitat de vegades que s'ha aplicat el mateix tipus de procediment dins el conjunt de procediments agrupats en un mateix número d'ingrés.

Els elements que apareixen dins de cercles de color vermell i verd a la Figura 3.8 són els procediments principals que es representaran en la visualització del graf resultant. Es pot veure doncs, que cadascun d'ells conté una etiqueta que indica a quin número d'ingrés pertany per tal d'agrupar-los segons el l'ingrés corresponent.

3.2.4.1 Utilitat en la investigació oncològica

Els resultats que s'obtenen en aquesta consulta permeten disposar de la quantitat de vegades que es repeteix el procés d'aplicar un mateix procediment principal durant els ingressos associats a les persones que van ser diagnosticades d'un tumor en concret en el passat.

A diferència dels resultats que s'obtenen a la primera consulta, la informació resultant presenta l'agrupació de les persones que se'ls hi va aplicar un mateix procediment principal en un mateix número d'ingrés, de manera que es visualitzen cronològicament segons l'ordre dels ingressos als quals pertanyen.

Gràcies a aquesta visualització, és possible estudiar detingudament el conjunt de persones que segueixen el mateix camí, pel que fa els procediments principals aplicats, des de que van ser diagnosticades d'un tumor en concret fins a l'últim ingrés. És a dir, permet avaluar els diferents procediments principals aplicats en aquests ingressos i com es relacionen entre ells cronològicament per tal d'extreure les regles i algorismes que seran introduïdes al software ASEDAT [3].

3.2.4.2 Representació dels resultats

Tal i com es pot veure a la Figura 3.9, la informació resultant es representa mitjançant un graf. Partint del node vermell que representa el total de les persones diagnosticades d'un tumor X, podem arribar als següents dos nodes de color negre. El primer d'aquests dos agrupa el conjunt de persones que van ingressar després de ser diagnosticades i, els segon, ho fa pels que no ho van fer mai. Continuant per les arestes que parteixen del node de color negre que representa el primer ingrés posterior al diagnòstic del tumor, hi podem veure els diferents tipus de procediments principals aplicats en aquests primers ingressos. Aquests nodes, es distingeixen gràcies al seu color taronja. De fet, per a fer més entenedor el resultat, aquests procediments queden agrupats en els nodes de color groc destinats a representar els procediments genèrics que contenen procediments principals de característiques similars.

A la Figura 3.9 es pot veure que, per a cada conjunt de persones que durant el seu primer ingrés se'ls hi va aplicar el mateix procediment principal representats mitjançant nodes grocs, parteixen dues arestes que es dirigeixen cap a dos nodes de color negre. Aquests dos, representen el total de persones que han re-ingressat després d'haver-los-hi aplicat el procediment indicat en el node groc en el seu primer ingrés. A partir d'aquí, la informació per a tots els posteriors re-ingressos, es representa de la mateixa manera que la que s'acaba de descriure pel primer i segon ingrés.

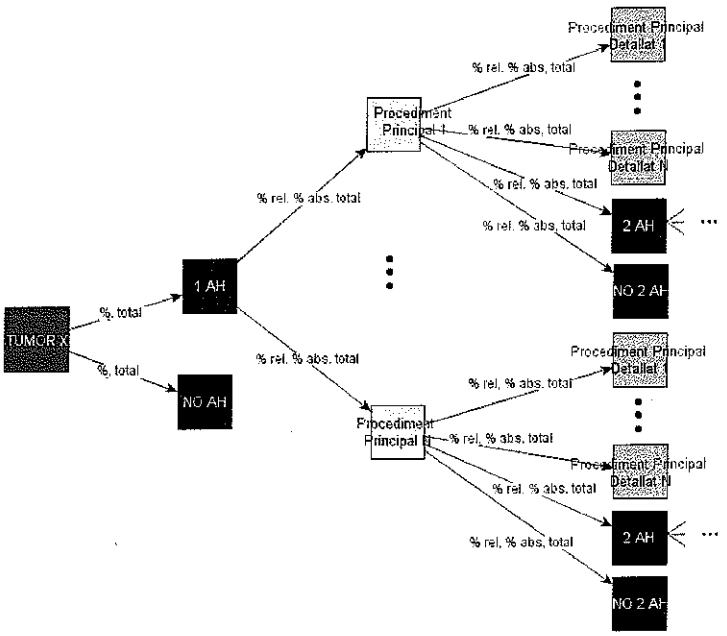


Figura 3.9: Visualització del graf resultant de la consulta 3

3.2.5 Consulta 4

Obté el conjunt dels diagnòstics principals i procediments principals que corresponen a la primera alta hospitalària dels individus que se'ls hi va localitzar un tumor determinat en algun moment de la seva vida.

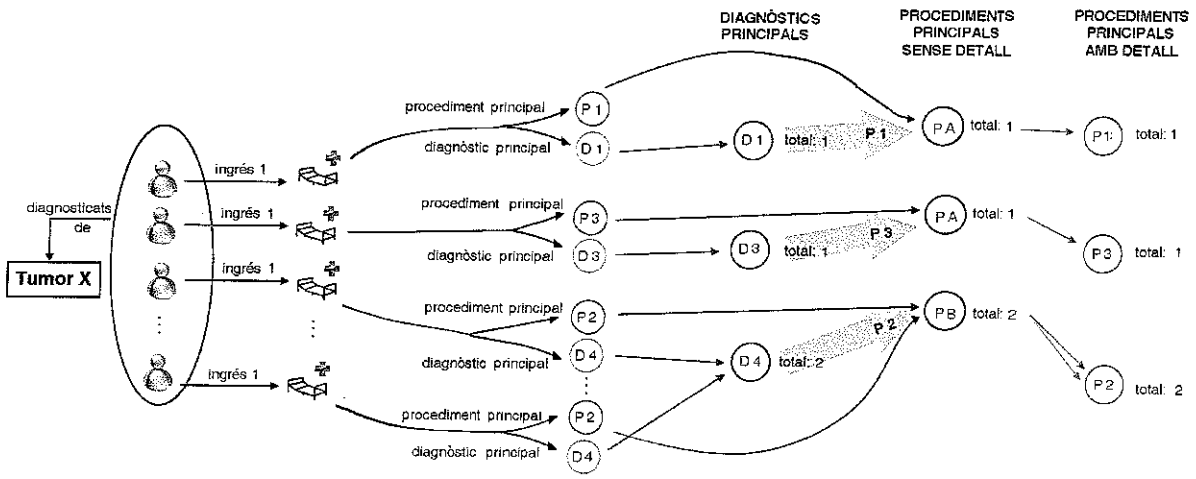


Figura 3.10: Definició de la consulta número 4

A diferència de l'anterior consulta, en aquest cas no ens interessa obtenir la relació entre els diferents ingressos hospitalaris de les persones involucrades. Tal i com es mostra a la Figura 3.10, partint de les persones diagnosticades d'un tumor X, cal obtenir les dades referents al primer ingrés posterior al diagnòstic d'aquest tumor. Per a cada primer ingrés, s'agafa el procediment principal i el diagnòstic principal que hi ha registrats. Un cop es disposa d'aquestes dades, és el moment de determinar els diferents tipus de diagnòstics principals que hi ha registrats dins aquest conjunt de primers ingressos.

A la Figura 3.10, es veu com hi ha un total de tres diagnòstics principals diferents. Cadascun d'aquests, conté el total de persones que tenen registrat el mateix diagnòstic principal durant el seu primer ingrés. El següent pas és determinar els diferents procediments principals que es van aplicar en els primers ingressos que han estat citats. De fet, ens interessa relacionar les persones que comparteixen un mateix diagnòstic principal i un mateix procediment principal.

A la Figura 3.10, es pot veure que els nodes rodejats d'un cercle de color vermell que representen els procediments genèrics. Aquests, agrupen les persones que comparteixen un mateix diagnòstic principal i un mateix procediment principal durant el seu primer ingrés. Tal i com s'ha explicat a la tercera consulta, els procediments genèrics permeten agrupar els procediments principals que tenen característiques molt semblants. Per tant, dins de cadascun d'aquests nodes representats mitjançant un cercle vermell, s'hi troben diferents procediments principals.

3.2.5.1 Utilitat en la investigació oncològica

Cal dir que la tercera, quarta i cinquena consulta estan destinades a permetre la total identificació i determinació de regles i algorismes per a ser introduïdes a l'ASEDAT [3] mitjançant l'avaluació dels resultats que ofereixen.

En aquesta en concret, la informació resultant permet analitzar quins són els diagnòstics principals registrats en els primers ingressos de cada persona, a més d'obtenir la freqüència d'aparició per a cada tipus de diagnòstic diferent. Per a cada conjunt de persones que comparteixen el mateix diagnòstic principal, es mostren tots els procediments principals que se'ls hi van aplicar durant el primer ingrés. La finalitat, altre cop, és disposar d'informació útil i garantir l'avaluació complerta dels procediments principals aplicats amb l'objectiu de facilitar la tasca d'extracció de regles i algorismes per a ser introduïdes al software esmentat.

3.2.5.2 Representació dels resultats

Tal i com es pot observar a la Figura 3.11, la representació dels resultats és força similar a la de la tercera consulta. La diferència més significativa és el fet que tan sols es representa la informació resultant pel primer ingrés i, per tant, el graf no té més continuïtat que la que es presenta en aquesta figura.

Com a totes les consultes, el node vermell agrupa el total de les persones diagnosticades d'un tumor en concret. A partir d'aquí, seguint per les dues arestes, hi trobem dos nodes de color negre que representen el primer ingrés d'aquestes persones, de la mateixa manera que ho fa la tercera consulta. Les arestes que parteixen d'aquests nodes de color negre, es dirigeixen als nodes de color taronja que agrupen els diferents diagnòstics principals registrats en els primers ingressos. Seguint per les arestes que parteixen d'aquests diagnòstics principals, s'arriba als nodes de color groc que representen els diferents procediments principals registrats durant el primer ingrés. Concretament, cadascun d'aquests nodes, agrupa el conjunt de persones que comparteixen un mateix diagnòstic principal i un mateix procediment principal durant el primer ingrés. De fet, els nodes de color groc són procediments genèrics que engloben als procediments principals de la mateixa manera que es fa a la tercera consulta.

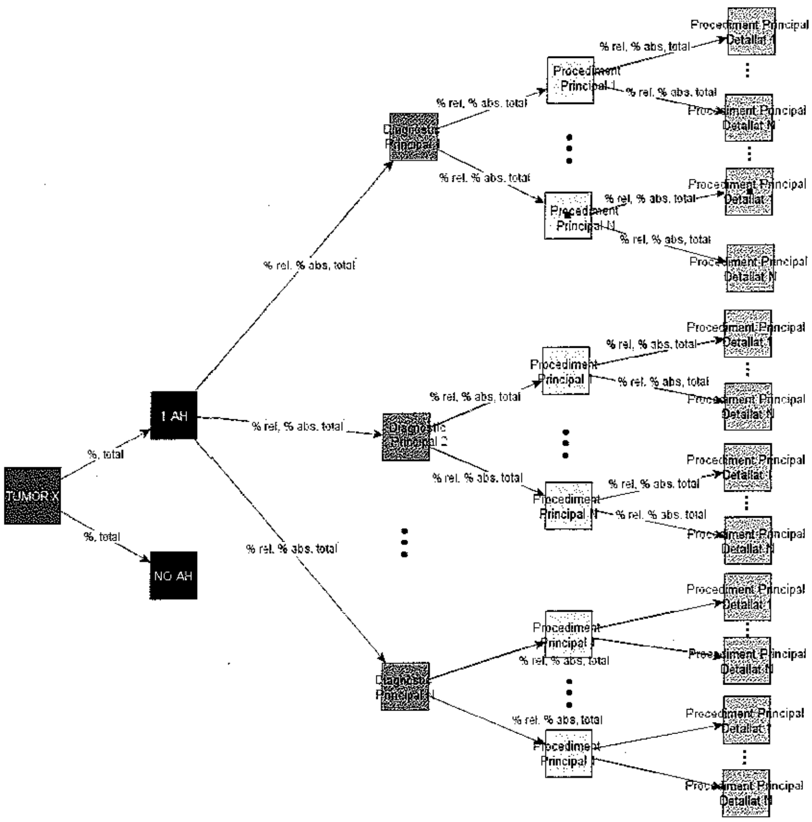


Figura 3.11: Visualització del graf resultant de la consulta 4

3.2.6 Consulta 5

Obté el conjunt de tots els procediments principals i diagnòstics principals d'entre totes les altes hospitalàries relacionades amb els individus que se'ls hi va localitzar un tumor determinat en algun moment de la seva vida, classificats pel número d'alta hospitalària i ordenats cronològicament.

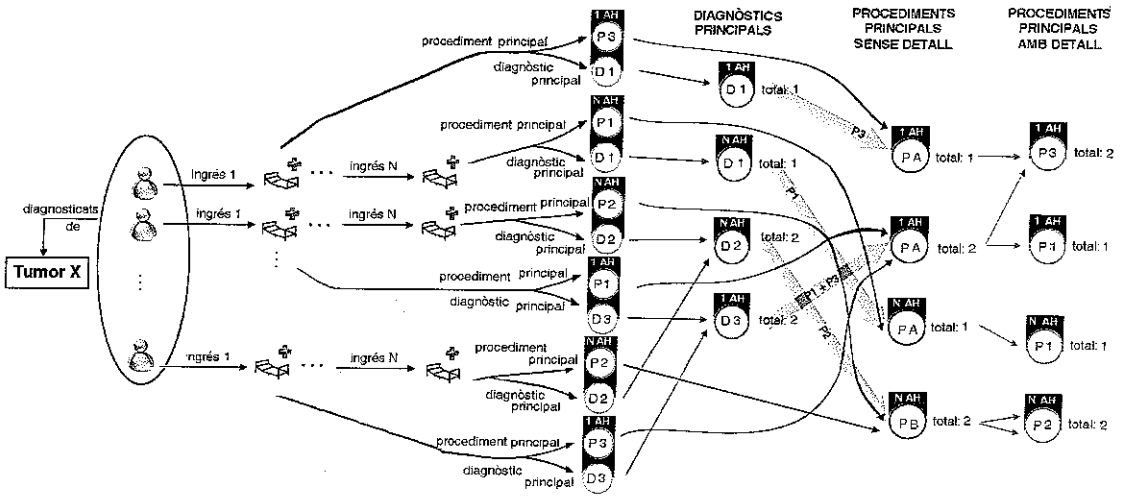


Figura 3.12: Definició de la consulta número 5

Els resultats que s'obtenen en aquesta consulta són pràcticament els mateixos que a la quarta consulta. La única diferència és el fet que, en l'anterior, tan sols es mostrava la informació per al primer ingrés de cada persona. En canvi, en aquesta, es representa la informació per a tots els ingressos i no només per als primers. De fet, els resultats que s'obtenen també són similars a la tercera consulta ja que comparteixen el fet que la informació es presenta seguint una cronologia d'ingressos que, en aquest cas, també s'hi representen els diagnòstics principals.

Tal i com es mostra a la Figura 3.12, es parteix del conjunt de persones diagnosticades d'un tumor en concret. Per a cada persona, es prenen les altes hospitalàries posteriors al diagnòstic d'aquest tumor i, per a cadascuna d'aquestes altes, s'obtenen el procediment principal i el diagnòstic principal associats. Un cop es disposa d'aquesta informació, és el moment de determinar els diferents diagnòstics principals registrats durant el primer ingrés, segon ingrés, etc. fins a l'últim ingrés de cada persona. Aquest procés es troba representat a la Figura 3.12 mitjançant les arestes i nodes de color lila. D'aquesta manera, s'obté un conjunt de diagnòstics diferents per a cada número d'ingrés. El següent pas és identificar els diferents procediments principals aplicats en aquells ingressos que tenen un mateix diagnòstic principal. La Figura 3.12, representa aquest procés mitjançant les arestes i els nodes de color vermell. Concretament, per a cada número d'ingrés, els nodes de color verd representen els diferents procediments principals registrats, de manera que queden agrupats dins un procediment genèric representat dins un node de color vermell.

L'objectiu d'aquesta consulta doncs, és obtenir la relació entre les persones que comparteixen un mateix diagnòstic principal en un número d'ingrés determinat i, alhora, veure els procediments principals que se'ls hi ha aplicat i com es relacionen entre ells.

3.2.6.1 Utilitat en la investigació oncològica

Tot i què els resultats que s'obtenen en aquesta consulta són pràcticament iguals que en la quarta, cal mencionar que en aquest cas, s'obté informació per a tots els ingressos i no es limiten a mostrar tan sols els del primer. Així doncs, els resultats són encara més útils per a avaluar els procediments diagnòstics i terapèutics que queden organitzats i ordenats en la visualització del graf resultant, i per tant, permet determinar les regles i algorismes que, definitivament, s'introduiran al software ASEDAT [3].

3.2.6.2 Representació dels resultats

Atès que la representació dels resultats és pràcticament igual que la que s'ha descrit a la tercera consulta, no es durà a terme una descripció extensa.

A la Figura 3.13 es pot observar que la única diferència es troba en el fet que, en aquest cas, existeix un nou conjunt de nodes de color negre que parteixen dels nodes de color taronja que representen els diferents diagnòstics principals. Per tant, aquests nodes negres agrupen el conjunt de persones que re-ingressen després d'haver ingressat per primer cop i que, a més, tenen registrat el mateix diagnòstic principal en aquest primer ingrés. A partir d'aquests nodes, la informació resultant és la mateixa que la que representen els elements que parteixen dels nodes de color negre del primer ingrés.

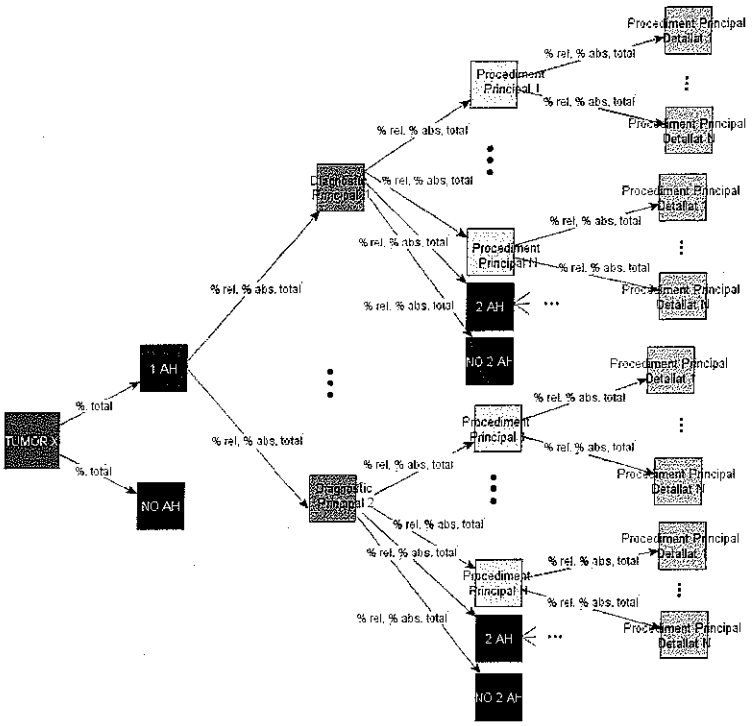


Figura 3.13: Visualització del graf resultant de la consulta 5

3.3 Flux de dades i informació

És important tenir clar quines són les dades d'entrada i quina informació genera l'aplicació com a resultat de la seva execució per tal d'entendre una mica més els objectius del projecte.

La Figura 3.14mostra quin és el flux de dades d'entrada i la informació de sortida.

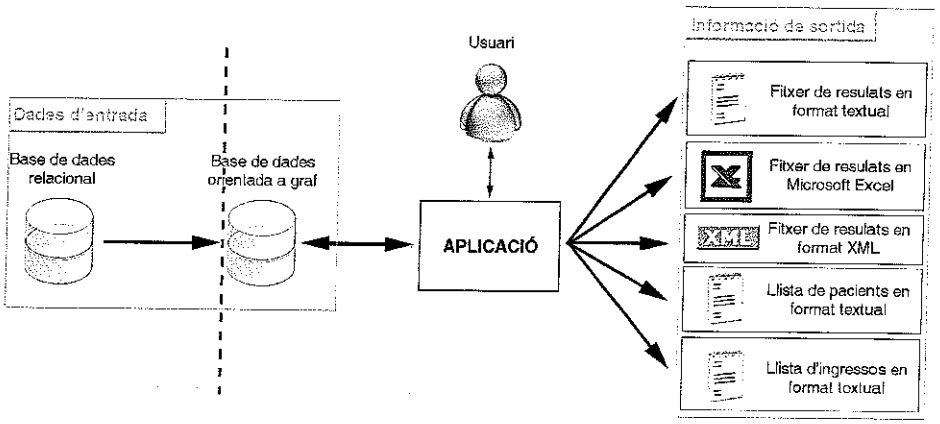


Figura 3.14: Flux de dades i informació durant l'execució de l'aplicació

Es pot veure doncs, com és necessària una Base de Dades relacional com a entrada de dades per a utilitzar l'aplicació. A més, el sistema software genera un total de cinc fitxers de sortida que contenen la informació resultant organitzada en diferents formats.

A continuació, es presenta la definició per a cadascun dels elements que componen el conjunt de les dades d'entrada. Per a més informació referent a quina és la presentació dels fitxers que contenen la informació de sortida cal consultar l'apèndix A a la pàgina 101.

3.3.1 Base de dades relacional

És l'element d'entrada imprescindible per a què el conjunt del sistema software tingui sentit. En el capítol anterior, han estat definits el RHT [25] i les AH de l'Hospital Universitari de Bellvitge i de l'ICO. Així doncs, la Base de Dades relacional sorgeix de la unió d'aquestes dues fonts de dades amb l'objectiu d'obtenir tots aquells pacients que van ser diagnosticats d'un únic tumor i que formen part del RHT juntament amb totes les seves altes hospitalàries posteriors a la data d'aquest diagnòstic que es troben a les AH.

La nova Base de Dades relacional reuneix les següents característiques:

- Conté un total de 50.000 tumors diagnosticats.
- Conté un total de 50.000 pacients.
- 1 tumor diagnosticat per a cada pacient de la Base de Dades.
- Conté un total de 90.000 altes hospitalàries.
- 2 altes hospitalàries aproximadament per a cada pacient de la Base de Dades.
- Dades registrades durant el període comprés entre 1994 i 2006.

El gràfic de la Figura 3.15 té com a objectiu exemplificar quines són les dades d'un pacient en cadascun d'aquests dos registres.

Podem veure com les dades de la mateixa persona que va ser diagnosticada d'un tumor determinat i es troben emmagatzemades al RHT també apareixen a la Base de Dades de les AH que conté tots els ingressos hospitalaris posteriors al diagnòstic del tumor.

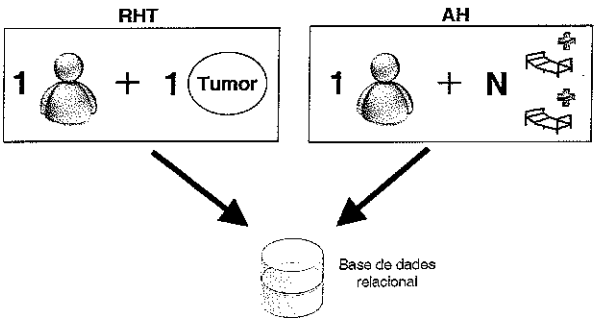


Figura 3.15: Exemple de les dades que fan referència a un pacient

Com es pot veure en l'esquema de flux de dades i informació de la Figura 3.14, la Base de Dades relacional està separada de la resta d'elements del gràfic mitjançant una línia discontínua de color vermell. Alhora, és possible veure que la fletxa que uneix aquesta Base de Dades i la Base de Dades orientada a graf indica, únicament, una transmissió de dades d'esquerra a dreta.

Aquesta distinció respecte a la resta es deu a què, en la majoria de les execucions de l'aplicació, no és necessari la càrrega de la Base de Dades relacional. Només en aquells casos que sigui el primer cop que s'inicia el programa o en el cas que s'actualitzi la Base de Dades relacional, es procedirà a dur a terme el moviment de dades en el sentit que indica la Figura 3.14. Per tant, en la majoria dels casos, el flux de dades i informació comença a partir de la línia discontínua.

Existeix informació més detallada referent a aquest tema en les seccions 4.2.2 a la pàgina 55 i 5.2.2 a la pàgina 70.

3.3.2 Base de dades orientada a graf

En aquest projecte es dóna la peculiaritat que la Base de Dades Relacional té una estructura molt similar a un graf.

Existeixen dues grans taules que emmagatzemen la informació referent als tumors diagnosticats i a les altes hospitalàries, respectivament. A més, cadascuna d'aquestes conté variables informatives per a completar la informació registrada per a cada tumor o alta hospitalària.

Així doncs, cada variable auxiliar és un identificador que es relaciona amb una altra taula. Aquestes variables poden representar hospitals, causes d'ingrés, procediments, diagnòstics, etc.

A la Figura 3.16, es mostra un esbós de l'estructura que segueix la Base de Dades relacional i com s'organitzen les relacions entre les diferents entitats.

Podem veure que, per tal què la Base de Dades tingui una estructura que permeti ser representada en un model orientat a graf, tan sols és necessari afegir una entitat auxiliar entre la taula de tumors diagnosticats i la d'altes hospitalàries amb l'objectiu de representar les persones.

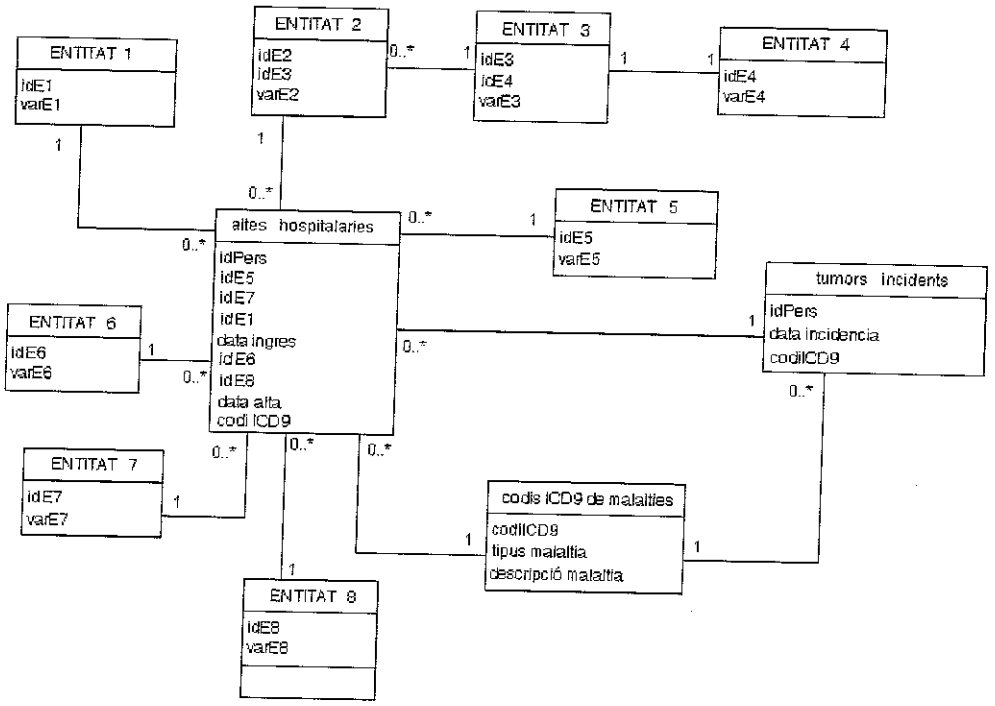


Figura 3.16: Estructura de la Base de Dades relacional

Cal tenir en compte que, en una Base de Dades relacional, les relacions entre entitats són implícites i tan sols es pot dur a terme la seva navegació mitjançant identificadors. En canvi, en una Base de Dades orientada a graf, no existeixen identificadors que representin claus foranes sinó que les relacions estan implícites en les arestes que uneixen les diferents entitats a relacionar.

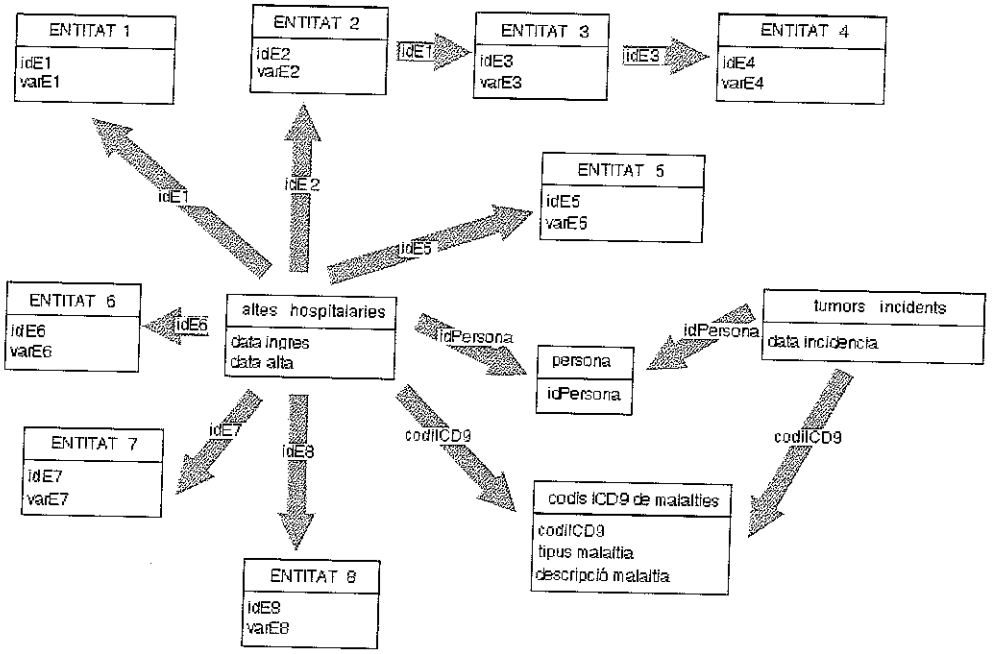


Figura 3.17: Estructura de la Base de Dades orientada a graf

La creació d'aquesta nova Base de Dades orientada a graf tan sols es realitzarà aquelles vegades que es carregui la Base de Dades relacional per primer cop. És a dir, un cop hagi estat construïda no caldrà

tornar-ho a fer. Ara bé, caldrà repetir el procés de construcció d'aquesta nova Base de Dades sempre que es modifiqui la Base de Dades relacional o sigui el primer cop que s'inicia l'aplicació.

Un vegada creada la Base de Dades orientada a graf, l'aplicació emmagatzema una imatge d'aquesta a disc amb l'objectiu de permetre'n l'accés durant l'execució del programa sense necessitat de construir-la cada vegada que sigui necessària l'obtenció de les seves dades. Aquesta creació es realitza mitjançant la llibreria *Jdex [6]* desenvolupada pel grup DAMA-UPC [14]. En aquest mateix capítol es presenta una definició detallada de *Jdex* a la secció 3.4.1. Per a més informació referent a aquesta Base de Dades tan sols cal consultar el capítol 5 a la pàgina 69.

3.4 Principal software a utilitzar

Cal fer menció dels recursos software que s'ha decidit utilitzar per tal de dur a terme la implementació d'aquest sistema software. El fet de treballar amb dades molt relacionades entre elles, tenir la intenció d'ampliar la Base de Dades relacional i planificar la creació de consultes complexes han estat els motius a tenir en compte per a seleccionar una eina o una altra.

3.4.1 Llibreria *Jdex*

Jdex [6] és un sistema de gestió de Bases de Dades orientades a graf creat pel grup DAMA-UPC [17].

3.4.1.1 Motivació

El motiu principal pel qual s'utilitza *Jdex* com a sistema gestor de la Base de Dades és per la gran quantitat de dades que cal tractar, la forta relació entre elles i la complexitat de les consultes a construir.

Tal i com s'ha definit prèviament, la Base de Dades en aquest projecte no és excessivament gran. Ara bé, es preveu que en un futur pròxim s'incorpori un volum considerable de dades que cal tenir molt en compte. En aquest cas, la manipulació de les dades mitjançant un sistema gestor de Bases de Dades relacional podria resultar un autèntic fracàs pel tipus de consultes a realitzar.

Cal mencionar també, la forta relació que mantenen les entitats de la Base de Dades. D'aquesta manera, si es pretén navegar a través de les dades contingudes en aquesta Base de Dades, és imprescindible la utilització d'un sistema gestor de Bases de Dades que en garanteixi l'eficiència i la rapidesa.

Consegüentment, *Jdex* ofereix les funcionalitats adients per tal de poder construir una aplicació que permeti executar consultes contra una Base de Dades de grans dimensions sense fracassar en rapidesa i comoditat. És per aquest motiu que *Jdex* serà l'encarregada de construir aquesta nova Base de Dades i gestionar-ne l'accés i manipulació de les dades.

3.4.1.2 Situació actual

La llibreria *Jdex* té el seu origen, com a primer prototip, en el desenvolupament d'una aplicació de reconeixement de patrons l'any 2006. Sorgeix de la necessitat d'emmagatzemar i gestionar fins a 25 GB de dades representades en una estructura en forma de graf que contenia més de 200 milions de nodes. Atesa la inexistència d'un sistema de gestió de Base de Dades orientat a graf eficient i capaç d'emmagatzemar grafs de grans dimensions, DAMA-UPC decideix crear-ne un de propi.

De seguida, sorgeix un nou projecte anomenat BIBEX [13]. Aquest doncs, és la primera interfície web que es desenvolupa mitjançant la nova llibreria *Jdex*.

DAMA-UPC decideix patentar-la el juliol de l'any 2008. Actualment, la llibreria resta pendent d'aprovació de la patent. Després de diferents proves, actualment *Jdex* suporta fins a 1.5 bilions de nodes i arestes, i és capaç d'emmagatzemar un total de 100 GB de dades.

Es preveu actualitzar-la a una nova versió que estarà disponible a finals de l'any 2008.

3.4.1.3 Característiques

Jdex reuneix un conjunt de característiques que la diferencien d'altres tipus de sistemes de gestió de Bases de Dades. És per aquest motiu, que és important conèixer-les.

Primer de tot, cal dir que *Jdex* permet representar gran quantitats de dades emmagatzemades en forma de graf. Les operacions que es realitzen sobre la Base de Dades són les mateixes que es poden realitzar sobre un graf. L'estructura d'emmagatzemament de dades en forma de graf permet definir etiquetes i afegir atributs per als nodes i arestes, establir direccions per a les arestes i, finalment, connectar dos nodes mitjançant més d'una aresta. Per tal de garantir l'emmagatzemament de grans volums de dades i la posterior recuperació d'aquestes, *Jdex* implementa estructures especials que ho fan possible.

Aquesta llibreria es caracteritza, principalment, per la seva eficiència en la gestió de grans volums de dades en termes de memòria. Quan es treballa amb Bases de Dades tan grans és molt important que el sistema que les gestiona en permeti un accés relativament ràpid. *Jdex* ofereix la possibilitat d'obtenir les dades pertinents sense necessitat de carregar tota l'estructura en forma de graf a memòria. A més, és possible navegar a través dels diferents elements que conformen el graf d'una manera ràpida i eficient. Per a accedir a una dada en concret, es disposa d'eines que eviten el recorregut exhaustiu del graf i permeten obtenir-la directament.

Jdex ofereix un ampli conjunt d'operacions per tal d'accedir a totes les dades de la Base de Dades orientada a graf. La diferència principal entre un sistema de gestió de Bases de Dades relacionals i un altre orientat a graf són les operacions d'aquest segon, les quals permeten navegar entre les relacions de les diferents entitats sense que el cost sigui un obstacle.

Aquest software ofereix suport per aquells desenvolupadors que necessiten accedir a dades fortament relacionades entre elles. La manipulació d'aquestes mitjançant un sistema de gestió de Base de Dades relacional pot resultar impossible atesa la gran relació entre les diferents entitats i, per tant, la constant navegabilitat a través d'aquestes. De fet, en Bases de Dades petites, el problema no es fa evident. Cal treballar amb grans volums de dades per a veure que és necessari utilitzar una Base de Dades orientada a graf per a navegar entre les diferents entitats.

Finalment, cal dir que *Jdex* permet crear noves estructures de graf amb l'objectiu d'emmagatzemar-hi la informació resultant generada durant una consulta a la Base de Dades.

3.4.1.4 Requisits tècnics

Per tal de garantir una correcta utilització del software, a continuació es presenten els requisits bàsics que ha de reunir el conjunt de dades, de software i de hardware:

- *Mida màxima del graf*: 50 vegades la memòria disponible
- *Número màxim de nodes*: 1.7 bilions per tipus de node
- *Número màxim d'arestes*: 1.7 bilions per tipus d'aresta
- *Número màxim de nodes i arestes*: 1.7 quintillons

- Microsoft Windows o Linux de 32 o 64 bits
- Java 1.5

3.4.2 L'eina de visualització *Prefuse*

Prefuse [24] és una llibreria de lliure distribució que ofereix un conjunt de funcionalitats per tal de crear visualitzacions de grafs interactius que permeten mostrar grans quantitats d'informació.

3.4.2.1 Motivació

El motiu principal pel qual s'ha optat per utilitzar *Prefuse* és el fet que, aquesta llibreria, permet visualitzar en forma de graf la informació obtinguda durant l'execució de les consultes. Cal recordar que aquesta informació ve representada en forma de graf gràcies a la llibreria *Jdex*, la qual cosa facilita la transformació de l'estructura de dades d'una llibreria a l'altra.

A més, permet representar la informació de la manera que s'havia previst per assegurar-ne una anàlisi clara i entenedora.

Finalment, cal fer menció especial al fet que es pot representar una part de la informació del graf de manera que tan sols cal indicar quina part es pretén explotar per a visualitzar-ne més informació.

3.4.2.2 Característiques

Prefuse és una de les llibreries més potents, de lliure distribució, que existeixen actualment per a visualitzacions de grafs. A continuació, s'exposen algunes de les principals característiques.

Aquesta llibreria suporta una gran quantitat de tipus de dades que, alhora, ofereix la possibilitat de crear una nova visualització i afegir-la dins la interfície gràfica d'una aplicació desenvolupada mitjançant el llenguatge de programació *Java* [18]. A més, permet construir visualitzacions animades amb el propòsit de donar moviment a la representació del graf.

Prefuse permet associar informació als elements del graf, crear-ne la visualització i garantir-ne la futura integració d'aquesta visualització dins de l'aplicació desenvolupada en *Java*. A més, disposa d'estructures d'emmagatzemament de dades com les taules, els grafs i els arbres depenent de la forma de la visualització que cal construir per tal de garantir una gestió de memòria més eficient. Gràcies a aquesta estructura de dades, permet construir vàries visualitzacions partint del mateix contingut.

Existeix la possibilitat de definir diferents estils per a cada grup d'elements que comparteixen les mateixes característiques. També es poden definir accions i associar-les a un d'aquests grups d'elements de la visualització amb l'objectiu de definir què ha de fer cadascun d'aquests elements en cada canvi que es produeixi. Finalment, també ofereix la possibilitat de definir *listeners* per tal que l'aplicació se n'assabenti de qualsevol canvi produït en la visualització.

Prefuse disposa d'eines molt útils i fàcils d'integrar a la visualització com són les cerques. També permet importar dades de diferents formats i integrar-les a la pròpia estructura de dades per a crear-ne la visualització.

Prefuse té la llicència BSD que pertany al grup de llicències de software lliure. La característica principal és que permet l'ús del codi font en un software que no sigui de lliure distribució. Ha estat implementada mitjançant *Java* així que es fàcilment integrable en aplicacions desenvolupades mitjançant el mateix llenguatge.

Capítol 4

Anàlisi i Disseny

En el desenvolupament d'una aplicació és necessari realitzar un seguiment del cicle de vida clàssic d'un sistema software.

Aquest capítol doncs, engloba la informació que fa referència a les primeres etapes d'aquest cicle que corresponen a l'anàlisi i disseny del sistema. Es defineixen els requisits funcionals i no funcionals del programa.

A més, també es presenta la descripció detallada del seu desenvolupament.

Cal afegir que s'utilitza l'estàndard UML per tal de definir l'anàlisi i disseny del sistema software atesa la seva llegibilitat i claredat, a més de la seva extensa utilització en el modelat de software actualment.

4.1 Anàlisi

En l'àmbit de l'enginyeria del software, l'anàlisi és una de les etapes més importants ja que és on s'aconsegueix definir quines són les condicions i capacitats requerides pel sistema i/o l'usuari.

Permet definir l'abast general per tal de detallar i documentar les funcionalitats que hauran de ser implementades en la futura codificació del software final.

L'etapa inicial d'anàlisi recull informació molt valuosa amb l'objectiu d'introduir-nos d'una manera més ràpida i concisa en el plantejament del disseny del software del producte i en la seva implementació, validació i manteniment.

Cada fase del cicle de vida clàssic ha de garantir el compliment dels requisits especificats en l'etapa d'anàlisi.

A la Figura 4.1 es mostra en detall quin és el procés de desenvolupament.

Cal tenir en compte que, en aquest projecte, es comença per l'anàlisi de requisits i es continua avançant per les etapes del cicle seguint el sentit que marca.

En el moment de finalitzar les proves, es procedirà a lliurar el sistema software al client. És a partir d'aquest instant que el client s'inicia en l'aprenentatge de la nova aplicació i informa al desenvolupador, si és necessari, dels defectes o millores que són necessàries efectuar.

D'aquesta manera, l'etapa de manteniment és l'encarregada de dur a terme aquesta última tasca, i sovint, cal tornar a realitzar totes les etapes del cicle de vida clàssic sempre que els canvis afectin a tot allò definit a l'especificació i disseny de l'aplicació.

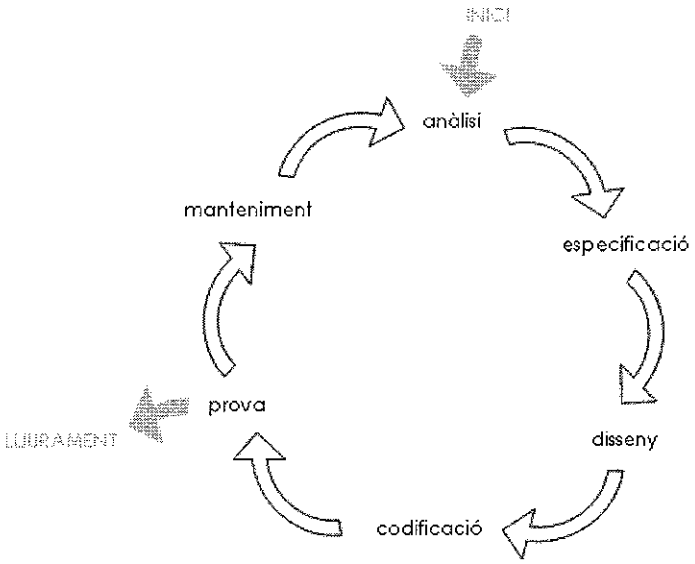


Figura 4.1: Cicle de vida

4.1.1 Anàlisi de requisits

L'anàlisi de requisits permet obtenir la llista de funcionalitats que l'usuari pretén veure-hi reflectides en l'aplicació software, així com la llista de condicions que ha de complir amb la finalitat de poder utilitzar-la correctament.

Cal destacar aquesta etapa del cicle pel fet que aquest projecte té com a objectiu desenvolupar una eina per a un client. Així doncs, han estat necessàries vàries reunions per tal de determinar i limitar les funcionalitats esperades per l'usuari final. Sense aquesta etapa inicial hauria estat totalment impossible començar a dissenyar la construcció del programa software.

4.1.1.1 Metodologia de l'anàlisi de requisits

Una de les problemàtiques del projecte en els seus inicis va ser el fet de no tenir clar quins eren els objectius de l'aplicació software a desenvolupar. Després de molt diàleg amb el client, va ser possible obtenir un esbós dels requisits del programa.

No obstant, per tal d'arribar a aquest punt va ser necessari el desenvolupament d'una eina exemple amb l'objectiu de mostrar les capacitats de les llibreries a utilitzar i quina seria la visualització dels resultats. L'exemplificació d'aquestes capacitats va permetre que la comunicació entre client i desenvolupador fos més fluïda i més concisa.

Finalment, i gràcies al disseny de la petita i senzilla eina exemple, es va poder definir detalladament la llista de funcionalitats definitiva per a la futura aplicació final.

4.1.1.2 Requisits funcionals

Els requisits funcionals són el conjunt de funcionalitats que l'usuari final espera poder trobar en el programa software. Aquestes funcionalitats poden ser classificades com explícites [E] si poden ser utilitzades directament per l'usuari, o implícites [I] si són d'utilització indirecta. En aquest projecte en concret, la llista de requisits funcionals és la següent:

- **[I] Tractament de la base de dades:** Oferir suport per tal de netejar, filtrar i eliminar tot tipus de soroll de les dades originals.
- **[I] Construcció de la nova base de dades:** Permet construir una nova base de dades orientada a graf, a partir de la base de dades relacional inicial.
- **[E] Gestió de projectes:** Crear, obrir, eliminar o tancar un projecte basat en una localització tumoral associada com a punt de sortida per tal de llançar totes les consultes que permet l'aplicació contra la base de dades carregada prèviament.
- **[E] Execució de consultes:** Realitzar una consulta determinada sobre la localització tumoral associada al projecte amb l'objectiu d'obtenir la informació desitjada segons la consulta elegida. Existeixen cinc tipus diferents de consultes que permeten l'estudi del comportament de cada tumor.
- **[I] Generació dels resultats en forma de graf:** Construir un graf amb la informació resultant d'una execució de les cinc possibles consultes en concret.
- **[I] Visualització del graf resultant:** Visualitzar, en forma de graf, la informació construïda damunt una estructura de graf en l'execució d'una de les cinc consultes.
- **[E] Manipulació de la visualització del graf:** Permet ampliar i reduir l'àrea de visió, navegar per tota la visualització del graf, desplaçar qualsevol node o aresta, seleccionar l'àrea de la visualització que es vol ampliar i visualitzar el graf sencer dins l'àrea de visió.
- **[E] Tipus de visualització del graf:** Permet elegir dos estils diferents de visualització del graf. Per una banda, és possible escollir la visualització de tota la informació del graf, i per l'altra, permet visualitzar un resum del graf que es pot anar expandint segons el node que interessa obtenir-ne més informació.
- **[E] Opcions de visualització del graf:** Modificar la distància en profunditat entre cada nivell de nodes, o bé, modificar la distància en amplada entre els nodes del mateix nivell.
- **[E] Visualització de la informació de cada element del graf:** Permet visualitzar en un requadre fix, per cada node i aresta de la visualització del graf, la informació continguda per cadascun d'aquests elements quan es passa el cursor del ratolí per damunt. En la mateixa acció, també apareix un nova finestra que mostra la mateixa informació d'una manera més clara i més detallada.
- **[E] Cerca en la visualització del graf:** Cerca i pinta en un color diferent de la resta de nodes del graf, tots aquells que coincideixen amb la informació introduïda.
- **[E] Generació de resultats en un fitxer de text:** Genera un fitxer de text amb tota la informació resultant de l'execució d'una consulta determinada. Així doncs, es representa la mateixa informació mostrada en la visualització del graf però organitzada de manera que es pot llegir en un fitxer de text.
- **[E] Visualització del resultats en un fitxer de text:** Permet visualitzar en una nova finestra de l'aplicació, la informació generada i organitzada en el fitxer de text.
- **[E] Cerca en el fitxer de text:** Cerca i selecciona aquelles paraules o fragments de paraules que coincideixen amb el text introduït a cercar.

- **[E] Generació de resultats en un full de càlcul:** Genera un full de càlcul amb la mateixa informació de la visualització del graf, però aquesta vegada, organitzada en una taula.
- **[E] Visualització del resultats en un full de càlcul:** Permet executar el programa Microsoft Excel per tal de visualitzar la informació generada prèviament i organitzada en una taula mitjançant un fitxer de full de càlcul.
- **[E] Generació de resultats en un fitxer XML:** Genera un fitxer XML amb la mateixa informació de la visualització del graf però en format XML, concretament, un format que permet visualitzar el mateix graf mitjançant altres aplicacions, i per tant, altres estils de visualització i manipulació.
- **[E] Generació de la llista de persones:** Genera un fitxer de text en el qual, per a cada node, existeix una llista d'identificadors de persones relacionades.
- **[E] Visualització de la llista de persones:** Permet visualitzar en una nova finestra de l'aplicació, l'identificador del node en concret i la llista d'identificadors de les persones que hi estan relacionades. Aquesta informació està estructurada en una taula.
- **[E] Cerca en la llista de persones:** Cerca, a partir de l'identificador de node introduït, la llista dels identificadors de persona que estan relacionats amb la informació que conté el node en concret.
- **[E] Generació de la llista d'altres hospitalàries:** Genera un fitxer de text en el qual, per a cada identificador de node, existeix una llista d'identificadors d'altes hospitalàries que hi estan relacionades.
- **[E] Visualització de la llista d'altes hospitalàries:** Permet visualitzar en una nova finestra de l'aplicació, l'identificador del node en concret i la llista d'identificadors de les altes hospitalàries que hi estan relacionades. Aquesta informació està estructurada en una taula.
- **[E] Cerca en la llista d'altres hospitalàries:** Cerca, a partir de l'identificador de node introduït, la llista dels identificadors d'alta hospitalària que estan relacionats amb la informació que conté el node en concret.
- **[E] Persistència:** Permet desar tots aquells resultats generats durant l'execució d'una consulta en concret; associant-los-hi la data de creació, la consulta que els ha generat, el projecte del qual formen part, els tipus de fitxer i el nom.
- **[E] Interfície gràfica:** Permet navegar a través de l'aplicació mitjançant un entorn gràfic suficientment còmode per tal de fer més amena la utilització del software.
- **[E] Ajuda:** Visualitza el manual d'instruccions del programa.

4.1.1.3 Requisits no funcionals

En aquest apartat, es llistaran el conjunt de requisits que no formen part de les funcionalitats de l'aplicació.

Interfície d'usuari i factors humans

L'usuari final que utilitzarà el sistema no ha de ser necessàriament un expert en l'àmbit de la informàtica sinó que pot interaccionar amb l'aplicació qualsevol tipus d'usuari genèric. No obstant, el fet que aquest software hagi de ser dissenyat amb l'objectiu de ser utilitzat per un col·lectiu d'investigació mèdica, fa que la interpretació dels resultats sigui pràcticament impossible sense tenir coneixements en l'àmbit de l'oncologia.

També cal dir que no existeix cap distinció entre els usuaris pel que fa a permisos d'accés. És evident que el programa software a desenvolupar ha de ser fàcil d'utilitzar i, per tant, molt intuïtiu ja que no es requereixen coneixements informàtics per a fer-ne ús.

A més, el fet d'utilitzar una Base de Dades relacional implementada sobre un conjunt d'arxius amb el format *Comma Separated Values* genera una llista de requisits importants per garantir un ús correcte de l'aplicació. És imprescindible que l'estructura i el format d'aquesta Base de Dades sigui tal i com s'especifica. Si no és així, aleshores és impossible la càrrega de les dades, i conseqüentment, el conjunt de funcionalitats incorporades al programa no es podrien utilitzar. Així doncs, serà necessari protegir el sistema de tots els possibles errors que puguin esdevenir. Per cada error, es mostrarà un missatge d'alerta i segons la seva importància i la part de l'aplicació que s'hagi vist afectada, es continuarà l'execució, o bé, es finalitzarà el programa.

Pel que fa al tipus de perifèrics d'entrada i sortida, tan sols és necessari disposar d'una pantalla, un teclat i un ratolí per tal d'interaccionar amb el programa software.

Documentació

Tota aplicació necessita una bona documentació per tal de fer més entenedor el conjunt de funcionalitats del sistema i, d'aquesta manera, garantir-ne el màxim aprofitament. L'aplicació constarà d'un manual d'usuari i una detallada descripció de la implementació del programa software i el seu respectiu estudi previ.

Consideracions de software i hardware

El sistema software que es pretén desenvolupar necessita ser executat sobre el sistema operatiu *Microsoft Windows* ja que un dels fitxers que es genera requereix ser visualitzat mitjançant l'eina *Microsoft Excel*. Per tal de veure el manual d'usuari és imprescindible tenir instal·lat un visualitzador de fitxers *Portable Document Format* (PDF). Tanmateix, existeix la possibilitat d'executar l'aplicació en un altre sistema operatiu com *Linux*, tenint en compte que no es podrà visualitzar la informació en un full de càlcul de *Microsoft Excel*, tot i què podrà ser generat i desat.

Encara que l'aplicació pot ser executada en un entorn hardware inferior al què és recomanable, cal tenir present quines són les característiques mínimes que permeten el correcte funcionament. Aquestes doncs, són les següents:

- Processador amb una freqüència igual o superior a 1 GHz.
- Memòria RAM igual o superior a 512 MB.
- Espai mínim de 300 MB al disc dur.

Característiques de rendiment

Existeix un requisit important pel què fa el volum de dades a tractar. Tenint en compte que les consultes es realitzen sobre una Base de Dades que correspon al RHT [25] de l'Hospital Universitari de Bellvitge i de l'ICO, podem afirmar que el volum de dades a manipular és considerable.

El problema doncs, resideix en el fet de visualitzar tanta informació. És per aquest motiu que és recomanable no utilitzar més de 500.000 altes hospitalàries i/o 100.000 tumors diagnosticats ja que la velocitat de construcció i manipulació de la visualització dels resultats pot resultar feixuga, encara que això no limita el fet que es pugui partir de quantitats superiors.

Partint d'una Base de Dades relacional amb un total de 90.000 altes hospitalàries i 50.000 tumors incidents es requereix, aproximadament, 5 minuts per a carregar-ne les dades i transformar-les a una nova Base de Dades orientada a graf. Aquesta estimació de temps s'ha dut a terme mitjançant un processador Intel amb freqüència d'1,5 GHz i 1 GB de memòria RAM. En el pitjor dels casos, l'execució d'una consulta sobre la Base de Dades orientada a graf requereix al voltant de 40 segons mitjançant el mateix maquinari descrit anteriorment.

Interfície del sistema

Per tal de construir la nova Base de Dades orientada a graf és necessària l'entrada d'una Base de Dades relacional, amb l'objectiu de carregar-ne les dades, i d'un fitxer de configuració XML per tal de crear i obrir nous projectes durant l'execució l'aplicació software.

Pel que fa als fitxers de sortida que es generen en l'execució de l'aplicació, cal esmentar la imatge generada en el procés de construcció de la nova Base de Dades orientada a graf.

Finalment, per a cada consultada executada, es poden generar i desar fitxers en diversos formats. Aquestos poden ser fitxers de text, fulls de càlcul de *Microsoft Excel*, o bé, fitxers XML. Tots ells representen la informació resultant de l'execució de les diferents consultes.

Factors de qualitat

El software a desenvolupar requereix que es compleixin tots els requisits no funcionals especificats en els punts anteriors per assegurar-ne la màxima fiabilitat.

La importància de la portabilitat pel que fa aquest software en concret no és important, ja que el conjunt d'usuaris finals sempre utilitzen el sistema operatiu *Microsoft Windows* i disposen del programari requerit per l'aplicació instal·lat en el seus ordinadors. Tot i així, el sistema software garanteix la portabilitat en altres sistemes operatius.

Cal remarcar la usabilitat del software a desenvolupar ja que l'usuari final no és necessàriament un expert en informàtica, i per tant, s'ha de garantir la comprensió i facilitat d'aprenentatge, a més a més, de disposar d'una interfície amigable.

També es necessari tenir en compte l'eficiència en l'execució de consultes gràcies a la càrrega de les dades en una Base de Dades orientada a graf, la qual permet navegar per les dades i extreure'n informació d'una manera més ràpida.

Entorn de desenvolupament

Per a poder desenvolupar còmodament tot el sistema s'ha decidit utilitzar l'aplicació Netbeans 6.1 [23], atesa la seva funcionalitat d'anàlisi de memòria durant l'execució de l'aplicació que permet implementar el software aportant-hi més eficiència.

4.2 Especificació

Un cop s'ha realitzat l'anàlisi de requisits, és necessari iniciar la segona etapa del cicle de vida clàssic en el desenvolupament de qualsevol aplicació. Després de definir les funcionalitats del sistema, és el moment de descriure rigorosament com ha de ser construït aquest software per tal de facilitar-ne la posterior codificació.

4.2.1 Model de casos d'ús

En l'especificació d'un sistema informàtic no hi pot faltar la descripció detallada de cada cas d'ús. Cada cas, proporciona un curs d'esdeveniments típic que representa la interacció entre el sistema i l'usuari amb la finalitat d'assolir els objectius plantejats. Per tant, són utilitzats per a definir una seqüència d'accions que van succeint entre el sistema i l'usuari en resposta d'un esdeveniment iniciat per qualsevol d'aquestos dos.

4.2.1.1 Diagrama de casos d'ús

Primer de tot i abans de descriure en detall cada cas d'ús, cal tenir una visió general del conjunt que formen i representar-ho en un diagrama per a veure com interaccionen amb l'usuari. En aquest projecte, atès que l'usuari final és un usuari genèric i que no existeixen permisos que facin una especial distinció entre diferents tipus d'actors, tan sols tindrem un actor principal que estarà involucrat en tots els casos d'ús del sistema.

Per a facilitar-ne la lectura del diagrama, s'ha decidit agrupar els casos d'ús en un nivell superior de casos d'ús genèrics. El resultat final es pot observar a la Figura 4.2.

Figura 4.2: Diagrama de casos d'ús

4.2.1.2 Descripció dels casos d'ús

Després d'haver donat una primera visió del conjunt de casos d'ús mitjançant un diagrama, és el moment de descriure'n detalladament la seqüència d'accions que es duen a terme entre l'actor principal i el sistema. S'ha decidit continuar agrupant els casos d'ús en diferents casos d'ús genèrics, que no es descriuran, ja que no formen part del conjunt sinó que són una eina d'agrupament. Cal dir que existeix un únic actor per la qual cosa s'omet el fet d'especificar en cada cas d'ús quin és l'actor i qui és l'iniciador ja que en tots els casos és sempre el mateix.

Gestió de projectes

Cas d'ús	Crear projecte
Resum	Crear un nou projecte associant-hi una localització tumoral
Curs típic d'esdeveniments	
Accions de l'actor	Accions del sistema
1. L'usuari prem l'opció *Crear projecte.*	
	2. El sistema mostra un diàleg amb un quadre de text on cal introduir-hi el nom del nou projecte.
3. L'usuari introdueix el nom que desitja assignar-li al nou projecte i prem l'opció *Accepta.*	
	4. El sistema recull el nom introduït i li assigna al nou projecte. Acte seguit, mostra un nou diàleg amb la llista de possibles localitzacions tumorals que se li poden assignar.
5. L'usuari elegeix la localització tumoral que desitja assignar al nou projecte i prem *Accepta.*	
	6. El sistema recull la localització tumoral escollida, li assigna al nou projecte, defineix la configuració i el desa a disc. Acte seguit, carrega les dades de la configuració i prepara l'aplicació per a poder executar consultes.
Curs alternatiu	
3. L'usuari prem el botó *Cancel·la* i s'acaba el cas d'ús.	
3. L'usuari no introdueix cap nom al quadre de text, prem el botó *Accepta* i el sistema mostra un missatge d'alerta indicant que cal escriure un nom. Torna al punt 3.	
5. L'usuari prem el botó *Cancel·la* i s'acaba el cas d'ús.	

Cas d'ús	Obrir projecte
Resum	Obrir un projecte existent que ja té una localització tumoral assignada
Curs típic d'esdeveniments	
Accions de l'actor	Accions del sistema
1. L'usuari prem l'opció *Obrir projecte.*	
	2. El sistema mostra un diàleg amb la llista dels noms dels projectes desats a disc.
3. L'usuari selecciona el projecte que desitja obrir i prem l'opció *Accepta.*	
	4. El sistema recull el nom del projecte a obrir, carrega la configuració del projecte i prepara l'aplicació.
Curs alternatiu	
3. L'usuari prem el botó *Cancel·la* i s'acaba el cas d'ús.	

Cas d'ús	Eliminar projecte
Resum	Eliminar un projecte existent que té una localització tumoral assignada
Curs típic d'esdeveniments	
Accions de l'actor	Accions del sistema
1. L'usuari prem l'opció *Eliminar projecte*	
	2. El sistema mostra un diàleg amb la llista dels noms dels projectes desats a disc.
3. L'usuari selecciona el projecte que desitja eliminar i prem l'opció *Accepta.*	
	4. El sistema recull el nom del projecte l'elimina del disc juntament amb tota la informació associada que s'havia desat amb anterioritat.
Curs alternatiu	
3. L'usuari prem el botó *Cancel·la* i s'acaba el cas d'ús.	

Cas d'ús	Tancar projecte
Resum	Tancar el projecte obert en aquell moment.
Curs típic d'esdeveniments	
Accions de l'actor	Accions del sistema
1. L'usuari prem l'opció *Tancar projecte*	
	2. El sistema elimina tota la informació creada temporalment associada al projecte i a les seves respectives consultes.

Executar consulta

El conjunt de casos d'ús que queden agrupats en aquest cas d'ús genèric tenen el mateix curs típic d'esdeveniments. Aquest fet es deu a què les cinc possibles consultes que es poden executar comparteixen el mateix diàleg. Cada consulta permet escollir quins fitxers addicionals, que representen el resultat de la consulta, han de ser generats a més a més del graf resultant i la seva visualització. Tot i què les consultes són diferents, el ventall de possibles fitxers a generar és el mateix per a cadascuna d'elles. És per aquest motiu, que tan sols es mostrarà el curs típic per a la primera consulta.

Cas d'ús	Consulta 1
Resum	Executa la consulta de cercar el conjunt de tots els procediments d'entre totes les altes hospitalàries relacionades amb els individus que se'ls va diagnosticar prèviament la localització tumoral del projecte obert.

Cas d'ús	Consulta 2
Resum	Executa la consulta de cercar el conjunt de tots els procediments d'entre totes les altes hospitalàries relacionades amb els individus que se'ls va diagnosticar prèviament la localització tumoral del projecte obert

Cas d'ús	Consulta 3
Resum	Executa la consulta de cercar quina és la cronologia del conjunt de tots els procediments principals d'entre totes les altes hospitalàries relacionades amb els individus que se'ls va diagnosticar prèviament la localització tumoral del projecte obert

Cas d'ús	Consulta 4
Resum	Executa la consulta de cercar el conjunt dels diagnòstics principals i procediments principals que corresponen a la primera alta hospitalària dels individus que se'ls hi va diagnosticar prèviament la localització tumoral del projecte obert

Cas d'ús	Consulta 5
Resum	Executa la consulta de cercar quina és la cronologia del conjunt dels diagnòstics principals i procediments principals d'entre totes les altes hospitalàries dels individus que se'ls hi va diagnosticar prèviament la localització tumoral del projecte obert

Curs típic d'esdeveniments	
Accions de l'actor	Accions del sistema
1. L'usuari prem l'opció *Consulta N* (sent *N* qualsevol de les 5 consultes).	
	2. El sistema mostra un diàleg amb una llista multi-opció que permet escollir quins fitxers addicionals han de ser generats.
3. L'usuari selecciona les opcions que cregui pertinents i prem el botó *Accepta*.	
	4. El sistema recull la llista de fitxers que ha de generar, executa la consulta contra la base de dades, genera el graf resultant i el visualitza, i finalment, genera els fitxers seleccionats.
Curs alternatiu	
3. L'usuari prem el botó *Cancel·la* i s'acaba el cas d'ús.	

Tipus de visualització

Cas d'ús	Visualitzar graf complert
Resum	Canvia la visualització del graf per tal que es visualitzi tota la informació resultant de la consulta executada anteriorment.

Curs típic d'esdeveniments	
Accions de l'actor	Accions del sistema
1. L'usuari prem l'opció *Veure graf complert*	
	2. El sistema genera una visualització del graf per a què es mostri tota la informació resultant de l'execució de la consulta.

Cas d'ús	Visualitzar graf resum
Resum	Canvia la visualització del graf per tal que es visualitzi una part de la informació resultant de la consulta executada anteriorment amb la possibilitat d'expandir aquells nodes que es desitgi obtenir més informació.

Curs típic d'esdeveniments	
Accions de l'actor	Accions del sistema
1. L'usuari prem l'opció *Veure graf resum.*	
	2. El sistema genera una visualització del graf per a què es mostri una part de la informació resultant de l'execució de la consulta.

Opcions

Cas d'ús	Opcions de filtratge
Resum	Mostra les opcions que permeten filtrar els nodes segons la quantitat de persones agrupades o el percentatge d'aquestes respecte altres nodes per tal de visualitzar el graf obviant allò que sigui irrellevant.

Curs típic d'esdeveniments	
Accions de l'actor	Accions del sistema
1. L'usuari prem l'opció *Opcions de filtratge.*	
	2. El sistema mostra un diàleg amb dues opcions que permeten elegir el filtratge per vegades repetides o per percentatge de repetició, i un quadre de text per introduir el valor de filtratge.
3. L'usuari selecciona l'opció que desitja i introdueix el valor de filtratge. Prem el botó *Accepta.*	
	4. El sistema recull l'opció elegida i el valor de filtratge associat i genera un nou graf resultant tot eliminant aquella informació que es repeteixi menys vegades que la indicada pel valor de filtratge. Acte seguit, visualitza el graf.

Curs alternatiu
3. L'usuari prem el botó *Cancel·la* i s'acaba el cas d'ús.
3. L'usuari no selecciona cap de les dues opcions i prem el botó *Accepta.* El sistema mostra un missatge d'alerta advertint a l'usuari que és necessari indicar l'opció de filtratge. Torna al punt 3.
3. L'usuari selecciona una de les dues opcions sense indicar el valor de filtratge, i prem el botó *Accepta.* El sistema mostra un missatge d'alerta advertint que és necessari indicar un valor. Torna al punt 3.
3. L'usuari selecciona l'opció de filtratge per percentatge de repetició, indica un valor fora de l'interval entre 0 i 100 i prem el botó *Accepta.* El sistema mostra un missatge d'alerta advertint que el valor de filtratge es troba fora de rang. Torna al punt 3.

Cas d'ús	Opcions de visualització
Resum	Mostra les opcions de visualització que permeten modificar la distància en profunditat entre els nodes de diferent nivell del graf i la distància en amplada entre els nodes del mateix nivell.

Curs típic d'esdeveniments	
Accions de l'actor	Accions del sistema
1. L'usuari prem l'opció *Opcions de visualització.*	
	2. El sistema mostra un diàleg amb dues barres de desplaçament que permeten modificar les dues distàncies mencionades.
3. L'usuari desplaça cada barra fins assolir el nivell que el satisfaci i tanca el diàleg.	4. El sistema recull, paral·lelament, els nous valors de les dues barres de desplaçament i modifica les distàncies pertinents en la visualització del graf.

Cerca informació

La cerca de persones relacionades a un node en concret i la cerca d'ingressos relacionats a un node comparteixen el mateix curs típic d'esdeveniments, per la qual cosa tan sols se'n definirà un de forma genèrica.

Cas d'ús	Cercar persones
Resum	Cercar els identificadors de persona relacionats amb la informació d'un node en concret.

Cas d'ús	Cercar ingressos
Resum	Cercar els identificadors d'ingrés relacionats amb la informació d'un node en concret.

Curs típic d'esdeveniments	
Accions de l'actor	Accions del sistema
1. L'usuari prem l'opció *Cercar {persones, ingressos}*	
	2. El sistema mostra una nova finestra on hi apareix una taula amb l'identificador del node a cercar i la llista dels seus identificadors relacionats.
3. L'usuari introdueix l'identificador de node a cercar i prem la tecla *Intro.*	
	4. El sistema recull l'identificador de node, cerca la llista d'identificadors associada i els mostra a la dreta de l'identificador de node. Torna al punt 3.
Curs alternatiu	
3. L'usuari introdueix un identificador de node que no existeix i prem *Intro.* El sistema mostra un missatge d'alerta indicant que no existeix. Torna al punt 3.	
3. L'usuari no introdueix cap identificador i tanca la finestra. S'acaba el cas d'ús.	

Cas d'ús	Cercar dins graf
Resum	Cercar quins nodes del graf resultant coincideixen amb la informació introduïda

Curs típic d'esdeveniments	
Accions de l'actor	Accions del sistema
1. L'usuari introdueix la informació a cercar dins el quadre de text de l'opció *Search* de la visualització del graf i prem *Intro.*	
	2. El sistema recull el text introduït i cerca els nodes que contenen aquest text. Acte seguit, si n'ha trobat, els dibuixa d'un color diferent en la visualització del graf.

Veure altres resultats

Cas d'ús	Veure resultats en text
Resum	Mostra la informació obtinguda en l'execució d'una consulta en format de text.
Curs típic d'esdeveniments	
Accions de l'actor	Accions del sistema
1. L'usuari prem l'opció *Format textual.*	
	2. El sistema mostra una nova finestra amb la informació resultant de l'execució de la consulta en format textual.

Cas d'ús	Veure resultats en full de càlcul
Resum	Mostra la informació obtinguda a partir de la consulta executada en un full de càlcul organitzat en taules.
Curs típic d'esdeveniments	
Accions de l'actor	Accions del sistema
1. L'usuari prem l'opció *Format full de càlcul.*	
	2. El sistema executa l'aplicació *Microsoft Excel* i mostra els fulls de càlcul generats prèviament.

Desar fitxers

Cas d'ús	Desar fitxers
Resum	Desar tots els fitxers generats en l'execució de la consulta a disc.
Curs típic d'esdeveniments	
Accions de l'actor	Accions del sistema
1. L'usuari prem l'opció *Desar fitxers.*	
	2. El sistema recull tots els fitxers temporals generats en l'execució de la consulta i els desa a disc junt amb el projecte obert.

Visualitzar informació node/aresta

Cas d'ús	Visualitzar informació node/aresta
Resum	Visualitza la informació que conté el node o aresta seleccionat per tal de mostrant-ne la mateixa informació o algun detall addicional.
Curs típic d'esdeveniments	
Accions de l'actor	Accions del sistema
1. L'usuari passa el cursor per damunt d'un node o aresta qualsevol de la visualització del graf.	
	2. El sistema recull l'identificador de node seleccionat i mostra un panell informatiu amb la informació detallada de l'element.

Mostrar ajuda

Cas d'ús	Mostrar ajuda
Resum	Mostra el manual d'usuari de l'aplicació
Curs típic d'esdeveniments	
Accions de l'actor	Accions del sistema
1. L'usuari prem l'opció *Mostrar ajuda.*	
	2. El sistema visualitza el manual d'usuari en una nova finestra mitjançant el visualitzador d'arxius PDF instal·lat.

Sortir del programa

Cas d'ús	Sortir del programa
Resum	Sortir del programa en execució
Curs típic d'esdeveniments	
Accions de l'actor	Accions del sistema
1. L'usuari prem l'opció *Sortir del programa.*	
	2. El sistema elimina tota la informació temporal creada durant l'execució de l'aplicació i es tanca el programa.

4.2.2 Model conceptual de la Base de Dades Relacional

Cal tenir en compte que en el disseny d'una Base de Dades relacional és important, com en qualsevol cas de l'enginyeria del software, tenir clars els requisits i objectius. Per aconseguir-ho, cal dur a terme una especificació que en defineixi aquests punts amb l'objectiu de facilitar-ne el futur disseny i realització de la Base de Dades.

4.2.2.1 Esquema conceptual

El disseny conceptual permet representar la Base de Dades sense haver de fer explícit el sistema de gestió de Base de Dades que serà necessari utilitzar. És evident que, després de dur a terme l'especificació, cal dissenyar l'esquema conceptual. L'especificació ha estat realitzada pel grup que desenvolupa el rol de client per la qual cosa no apareix en aquesta documentació.

A la Figura 4.3 es presenta l'esquema conceptual de la Base de Dades relacional de la qual parteix el sistema software a implementar.

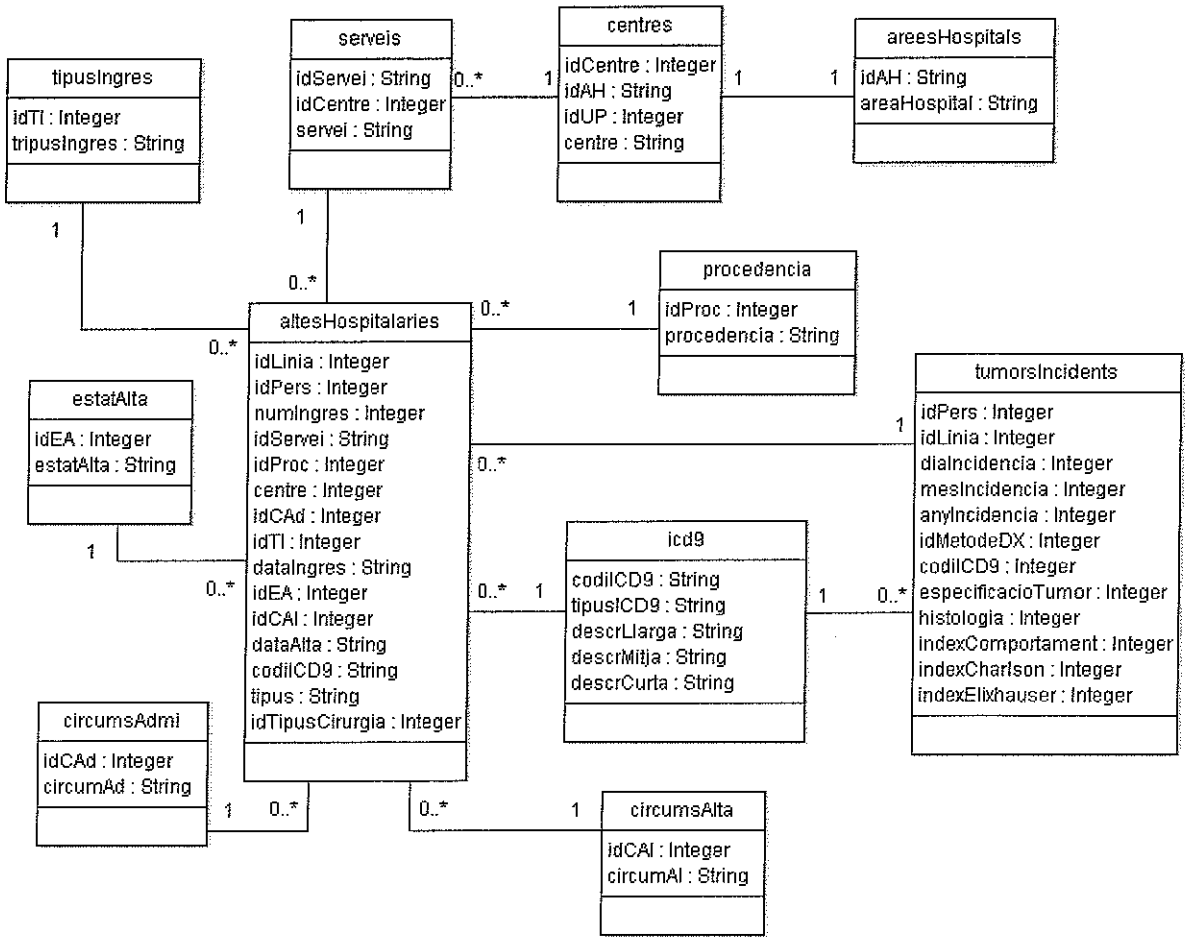


Figura 4.3: Esquema conceptual de la Base de Dades

4.2.2.2 Esquema lògic

Un cop s'ha definit l'esquema conceptual de la Base de Dades que permet al desenvolupador tenir clar l'estructura d'aquesta, és el moment de dissenyar l'esquema lògic a partir de la informació obtinguda en l'anterior esquema.

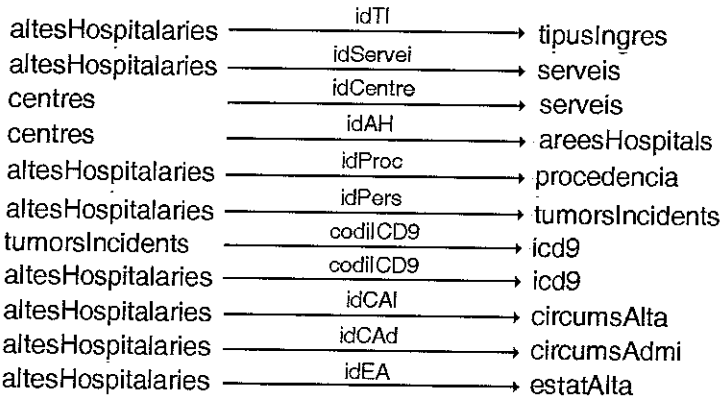
El disseny lògic és utilitzat per a definir més detalladament com serà l'estructura de la Base de Dades des del punt de vista informàtic. L'esquema lògic defineix el tipus de Base de Dades i el conjunt de relacions normalitzades. En aquest cas, es tracta d'una Base de Dades relacional.

A continuació es presenta l'esquema lògic de la Base de Dades amb les respectives relacions.

```
icd9(codiICD9, tipusICD9, descrLlarga, descrMitja, descrCurta)
```

```
circumsAdmi(idCAd, circumAd)
circumsAlta(idCAl, circumAl)
centres(idCentre, idAH, idUP, centre)
tipus_ingres(idTI, tipusIngres)
serveis(idServei, idCentre, servei)
procedencia(idProc, procedencia)
estatAlta(idEA, estatAlta)
areesHospitals(idAH, areaHospital)
altesHospitalaries(idLinia, idPers, numIngres, idServei, idProc, centre, idCAd, idTI,
                   dataIngres, idEA, idCAl, dataAlta, codiICD9, tipus, idTipusCirurgia)
tumorsIncidents(idPers, idLinia, diaIncidencia, mesIncidencia, anyIncidencia,
                idMetodeDX, codiICD9, especificacioTumor, histologia,
                indexComportament, indexCharlson, indexElixhauser)
```

En aquesta esquema, els noms de les relacions apareixen seguits dels noms dels seus respectius atributs entre parèntesi. Les claus primàries són els atributs subratllats amb una línia contínua i les claus foranes es representen mitjançant els següents diagrames referencials.



En aquest diagrama referencial podem veure que els noms de les taules apareixen a la columna de l'esquerra i a la columna de la dreta. L'esquema simbolitza que les taules contingudes a la primera columna estan relacionades mitjançant la clau forana indicada damunt de la relació que les uneix a les taules de la dreta.

4.2.3 Model conceptual del sistema software

El model conceptual d'un sistema software permet dibuixar quina serà l'estructura de les classes i com es relacionen entre elles, a més de definir-ne les operacions i els atributs principals que les conformaran. En la majoria dels casos, el model conceptual és d'una mida considerable atesa l'envergadura dels projectes a construir. Així doncs, es definirà un model conceptual per a cada mòdul de l'aplicació.

Cal tenir en compte que, cada mòdul, conté un conjunt de classes que comparteixen un elevat nombre de dades i, alhora, estan destinades a constituir funcionalitats de forma conjunta. No obstant, és evident que també existiran relacions entre classes d'altres mòduls de manera que, les classes que apareguin representades dins d'un mòdul i formin part d'un altre, seran etiquetades amb el nom del qual provenen i seran només especificades dins del mòdul corresponent.

Alhora, s'ha cregut convenient simplificar el detall de les operacions de manera que s'obviaran els paràmetres d'entrada i sortida ja que els noms són suficientment explícits. De totes maneres, en alguns casos és important l'especificació d'aquests paràmetres així que, en algunes classes del model, s'hi pot trobar el detall de les operacions.

Finalment, per tal de fer més intuïtiu l'esquema, tan sols es mostren aquelles operacions i atributs realment imprescindibles per a una correcta comprensió del model.

Mòdul *Image*

Aquest mòdul representa el conjunt de classes que gestionen la Base de Dades emmagatzemada a disc ja sigui per netejar-la, tractar-la o transformar-la al nou model de dades orientat a graf.

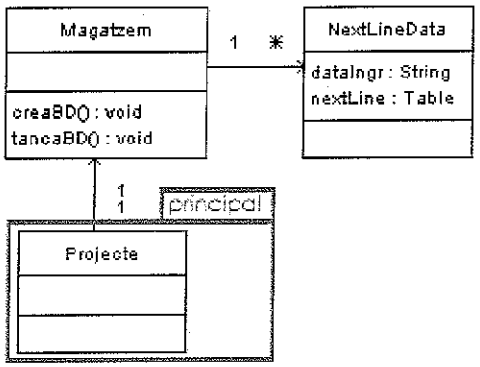


Figura 4.4: Model conceptual del mòdul *Image*

Mòdul *Utils Ui*

En aquest mòdul hi trobem les classes que són útils per a la completa realització de les operacions d'altres classes.

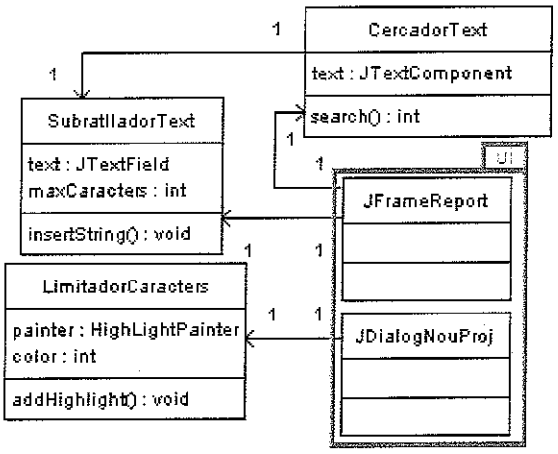


Figura 4.5: Model conceptual del mòdul *Utils Ui*

Mòdul *Principal*

El mòdul *Principal* forma part de tot allò que fa referència al domini d'una aplicació. Conté el conjunt de classes més importants ja que s'encarrega d'executar les consultes, gestionar els projectes i preparar les dades resultants de les execucions per tal de ser visualitzades.

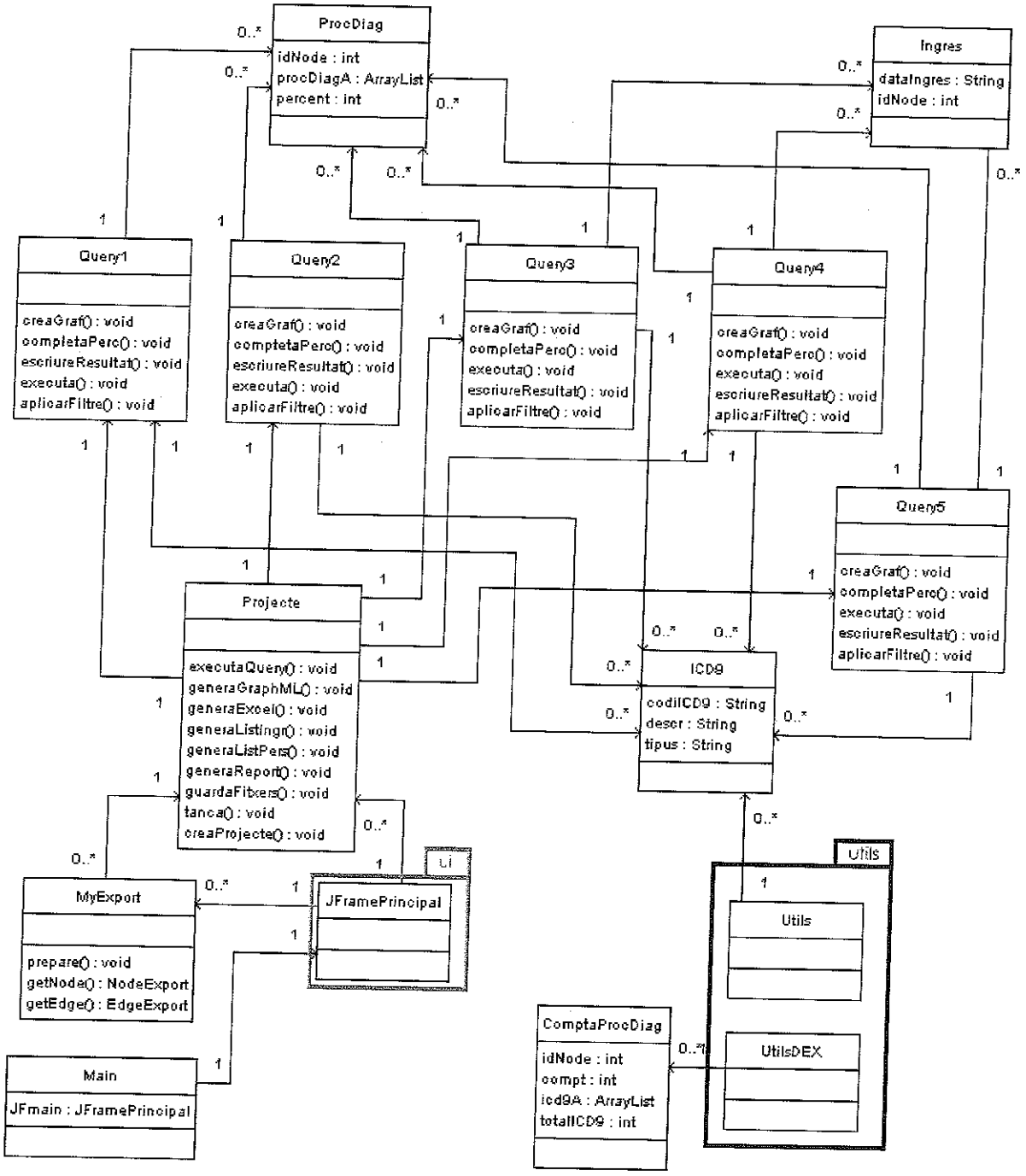


Figura 4.6: Model conceptual del mòdul *Principal*

Mòdul *Utils*

Les classes que componen aquest mòdul s'encarreguen de subministrar operacions útils per a la realització de les diferents operacions de la resta de classes del sistema. Tan sols es mostraran les classes que el componen junt amb les seves operacions i atributs, en aquest cas, de manera més rigorosa. No es mostren les classes amb les què es relaciona ja que són, pràcticament, totes.

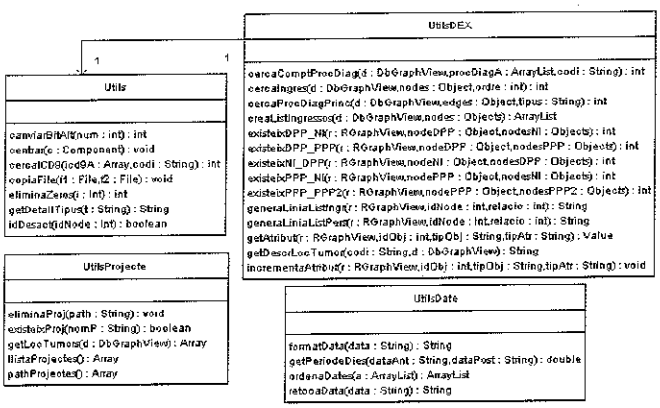


Figura 4.7: Model conceptual del mòdul *Utils*

Mòdul *Ui*

El conjunt de classes representat en aquest mòdul està directament relacionat amb tot allò lligat a la interfície gràfica i la visualització dels resultats.

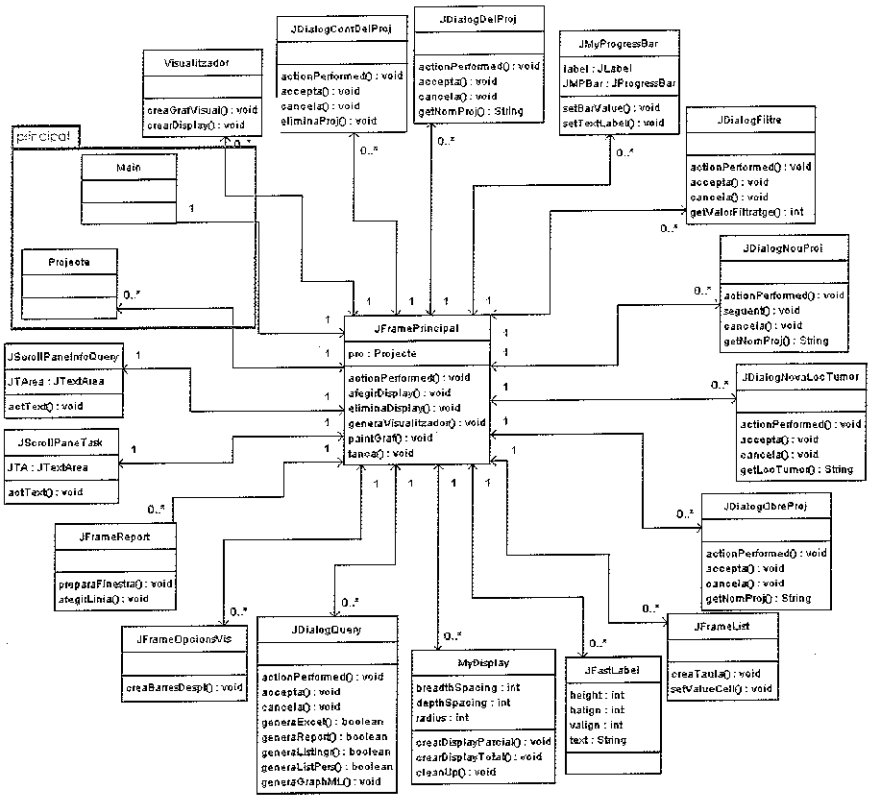


Figura 4.8: Model conceptual del mòdul *Ui*

4.3 Disseny

Amb l'anàlisi ha estat possible definir quin sistema calia construir. Després, amb l'especificació s'ha determinat què ha de fer l'aplicació software i amb quins recursos pot comptar, i finalment, amb el disseny es descriurà com ho ha de fer el sistema. El disseny és la fase que assegura una futura implementació correcta i ràpida.

4.3.1 Arquitectura

Tenint en compte que l'arquitectura és la definició del conjunt de subsistemes i components d'un sistema, així com la relació entre aquests, s'ha optat per desenvolupar un sistema software mitjançant una **arquitectura en 3 capes**.

El motiu pel qual s'ha elegit aquesta arquitectura és pel fet que garanteix una major canviabilitat, eficiència i usabilitat. D'aquesta manera, si en un futur es desitja canviar la visualització del graf mitjançant una llibreria més eficient i/o visual, l'arquitectura en 3 capes permet canviar la implementació ràpidament tan sols accedint a la codificació corresponent a la capa de presentació. De la mateixa manera, es podria canviar el que fa referència a la càrrega de les dades (capa de dades) i es podrien afegir o eliminar consultes contra la Base de Dades (capa de domini) sense que es vegi alterat la resta del sistema software.

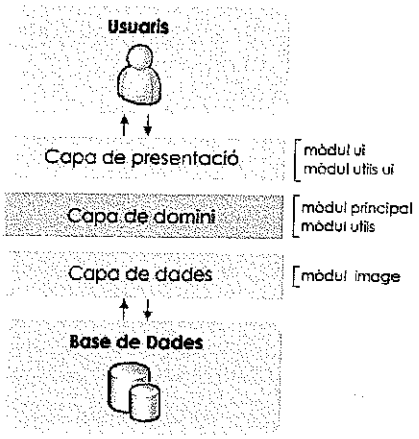


Figura 4.9: Arquitectura en tres capes

4.3.2 Patró arquitectònic

Un patró arquitectònic defineix un esquema d'organització de l'estructura d'un sistema software. En sistemes software de grans dimensions, és necessari tenir clar quin patró arquitectònic s'utilitza per tal de garantir-ne una organització correcta.

En aquest projecte en concret, s'ha optat per utilitzar un patró arquitectònic de **Crida i Retorn.** Aquest tipus de patró organitza l'estructura fonamental d'un sistema mitjançant un programa principal i un conjunt de components distribuïts en una arquitectura en capes tal i com s'ha definit prèviament. La comunicació entre el programa principal i els components es realitza jeràrquicament. És per aquest motiu que rep el nom de *Crida i Retorn*. A més, aquesta comunicació entre components es realitza mitjançant l'orientació a objectes, la qual cosa implica que cada classe del sistema agrupa un conjunt de dades i operacions a manipular. A la Figura 4.10 s'hi observa quina estructura segueix i com funciona.

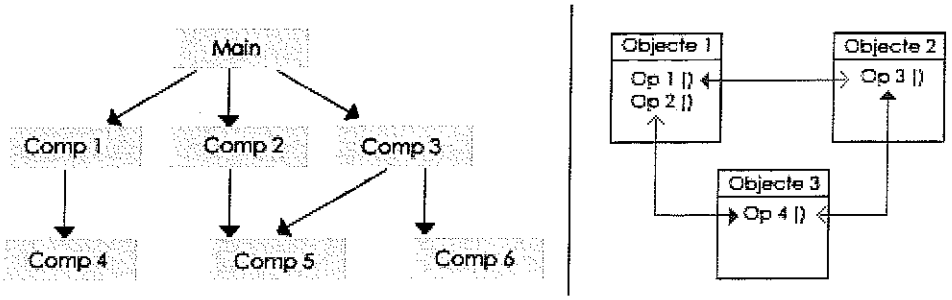


Figura 4.10: Estructura d'un patró *Crida i Retorn*

4.3.3 Model de comportament. Diagrames de seqüència

Les descripcions realitzades en el model de casos d'ús són un exemple clar de definicions amb l'objectiu de realitzar les últimes tasques del cicle de vida clàssic. Per cada cas d'ús rellevant és necessari definir el diagrama de seqüència de totes les operacions i dades que hi intervenen.

Per tal de simplificar aquest model, tan sols es representaran els diagrames de seqüència d'aquells casos d'ús que realment sigui interessant observar-ne el detall. Alhora, es classificaran els diagrames segons el cas d'ús genèric al qual pertanyen tal i com ja s'ha fet amb el model de casos d'ús.

Finalment, cal afegir que els diagrames de seqüència no representen un disseny exhaustiu sinó que són una visió molt aproximada del que serà el resultat final de la implementació. Totes aquelles classes de la interfície gràfica que intervenen en les diverses seqüències queden agrupades en la classe principal anomenada *JFramePrincipal* amb l'objectiu de simplificar-ne els diagrames.

Gestió de projectes

Pel que fa els casos d'ús de *Crear projecte* i *Obrir projecte*, podem afirmar que comparteixen un mateix diagrama de seqüències. L'única diferència és que, en el primer cas, l'usuari indica el nom del projecte i la localització tumoral associada, i en el segon cas, tan sols ho fa amb el nom del projecte per a què el sistema el pugui localitzar i prendre la localització tumoral que té associada. Així doncs, es representarà qualsevol dels dos casos d'ús.

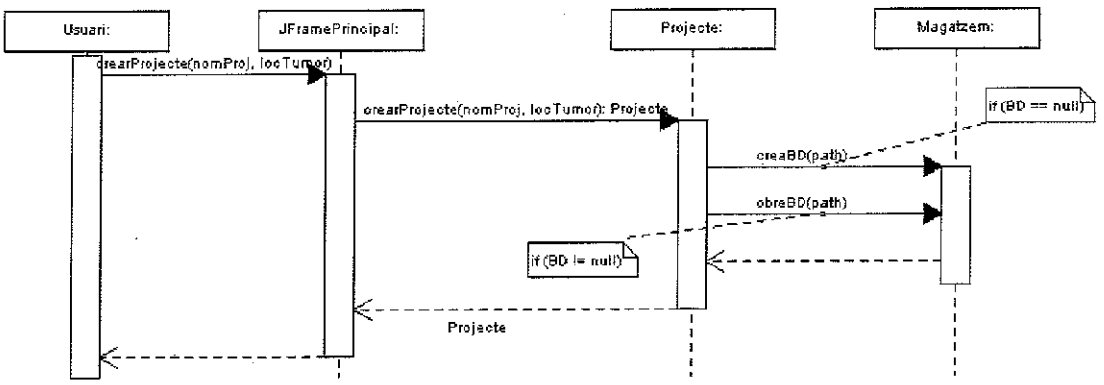


Figura 4.11: Diagrama de seqüència de *Crear projecte*

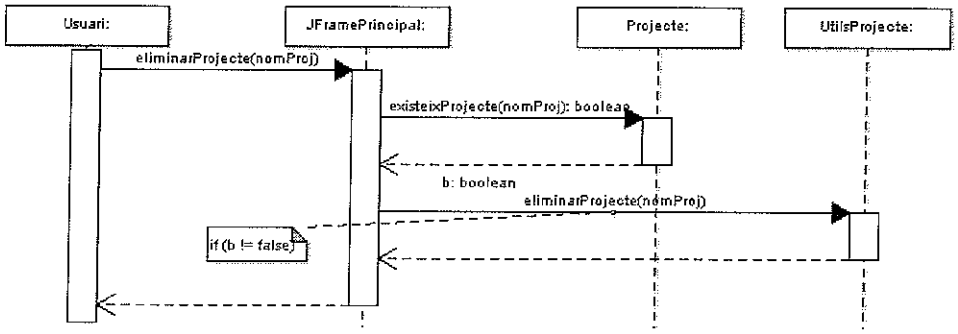


Figura 4.12: Diagrama de seqüència de *Eliminar projecte*

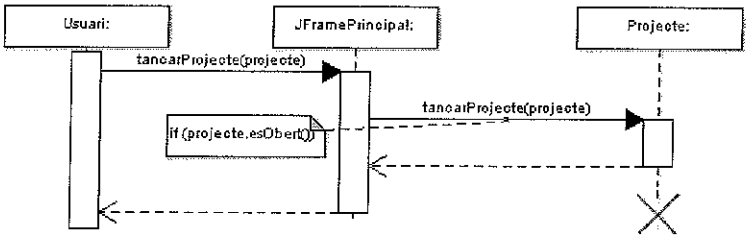


Figura 4.13: Diagrama de seqüència de *Tancar projecte*

Desar fitxers

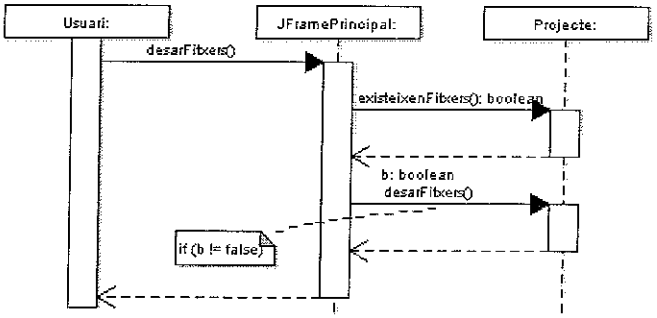


Figura 4.14: Diagrama de seqüència *Desar fitxers*

Executar consulta

El cas d'ús genèric *Executar consulta*, engloba les seqüències que fan referència a les cinc consultes que constitueixen l'aplicació. Tot i què els resultats que s'obtenen en cada consulta són diferents, el diagrama de seqüències és exactament el mateix per a totes elles degut a la gran similitud del procés d'execució per a cada consulta. Les operacions comparteixen el mateix nom per a totes les consultes ja que l'objectiu en cada consulta és el mateix. Com a exemple, es definirà el diagrama de la primera consulta.

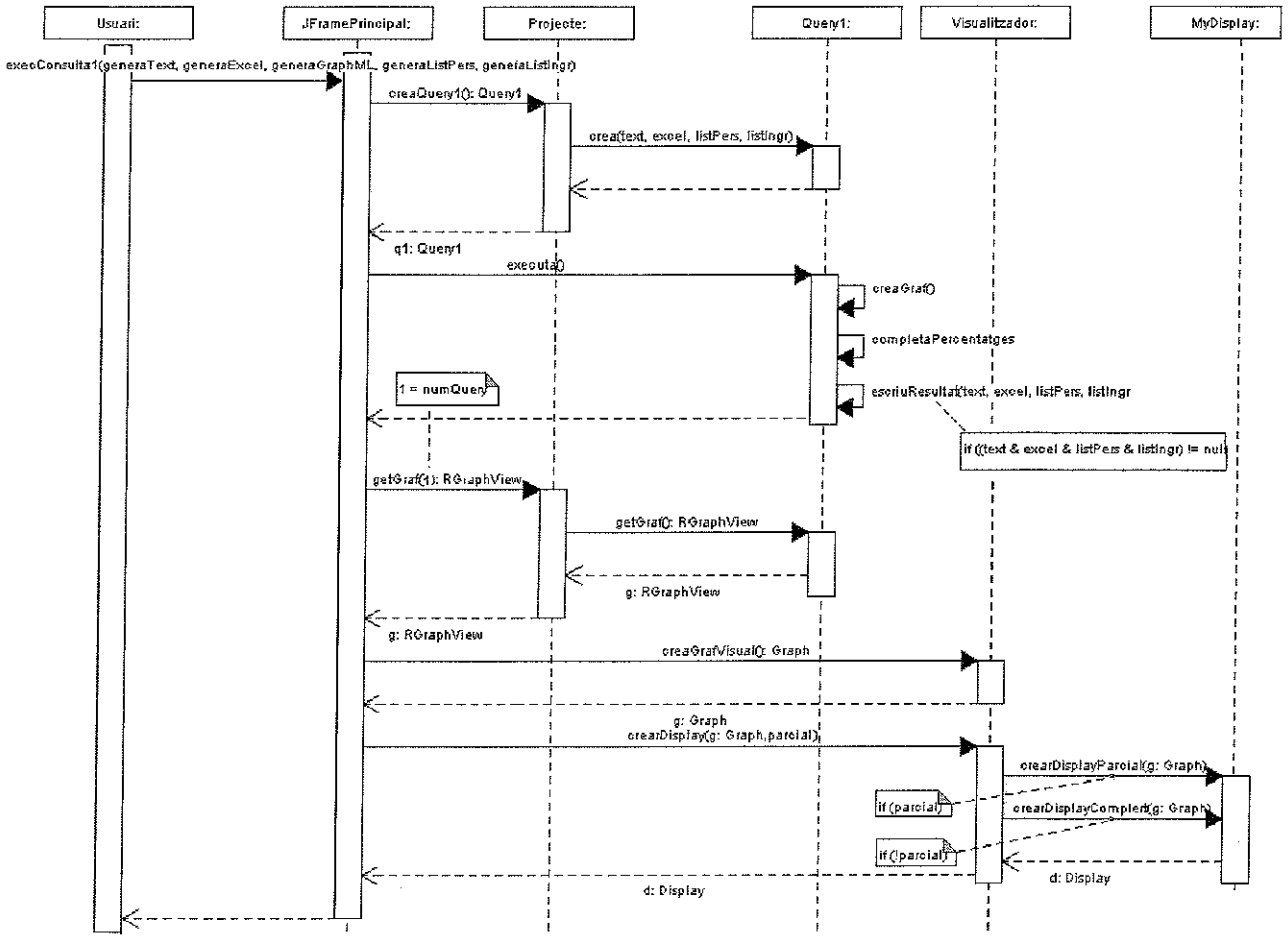


Figura 4.15: Diagrama de seqüència *Executar Consulta*

Veure altres resultats

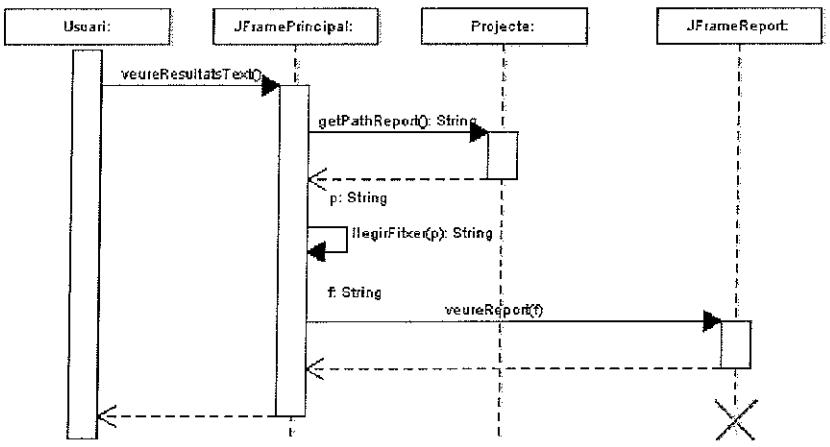


Figura 4.16: Diagrama de seqüència *Veure resultats text*

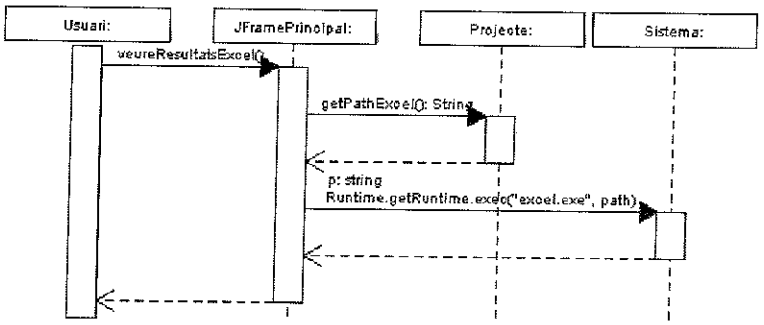


Figura 4.17: Diagrama de seqüència *Veure resultats excel*

Tipus de visualització

En aquest cas d'ús genèric hi apareixien els casos d'ús de *Graf Complert* i *Graf Resum.* Tan sols es definirà el cas *Graf Complert* ja que el diagrama de seqüència d'ambdós casos es pràcticament el mateix. L'única diferència és en la crida de l'operació *crearDisplay(g: Graph, parcial: boolean)*. D'aquesta manera, quan es vulgui visualitzar el graf complert amb tota la informació caldrà passar com a paràmetre el *graf visual* i el booleà a *false* per tal de crear un display que visualitzi el graf complert. En canvi, si es desitja visualitzar el graf resum, la crida es realitzarà mitjançant els paràmetres *graf visual* i el booleà a *cert* amb l'objectiu de crear un display amb el graf resum.

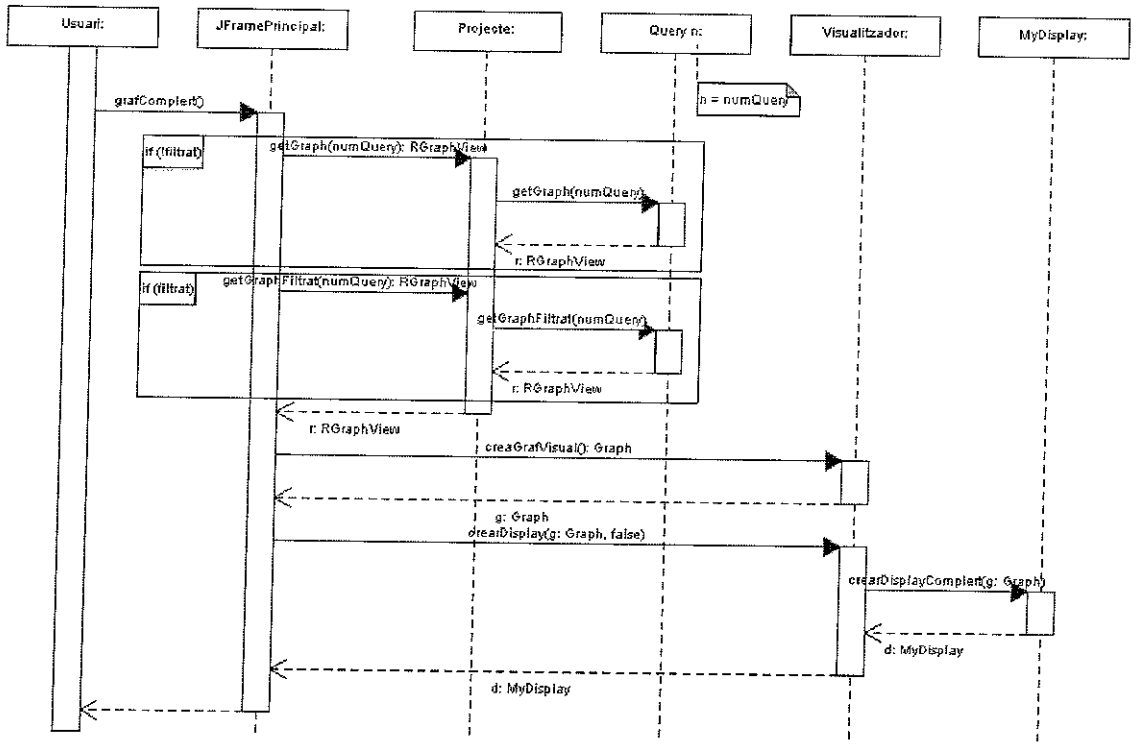


Figura 4.18: Diagrama de seqüència *Graf Complert*

Opcions

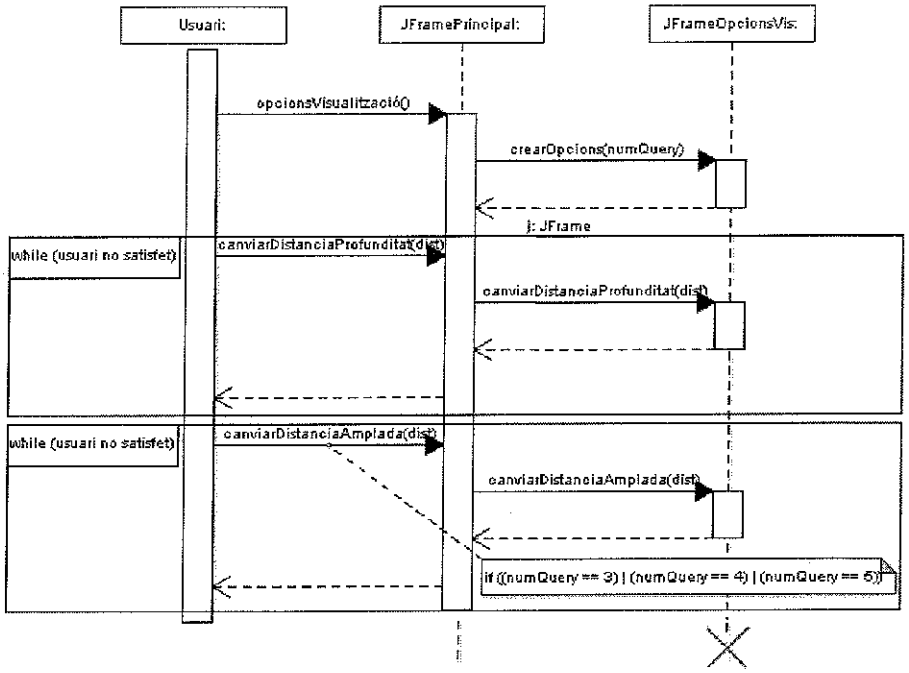


Figura 4.19: Diagrama de seqüència *Opcions Visualització*

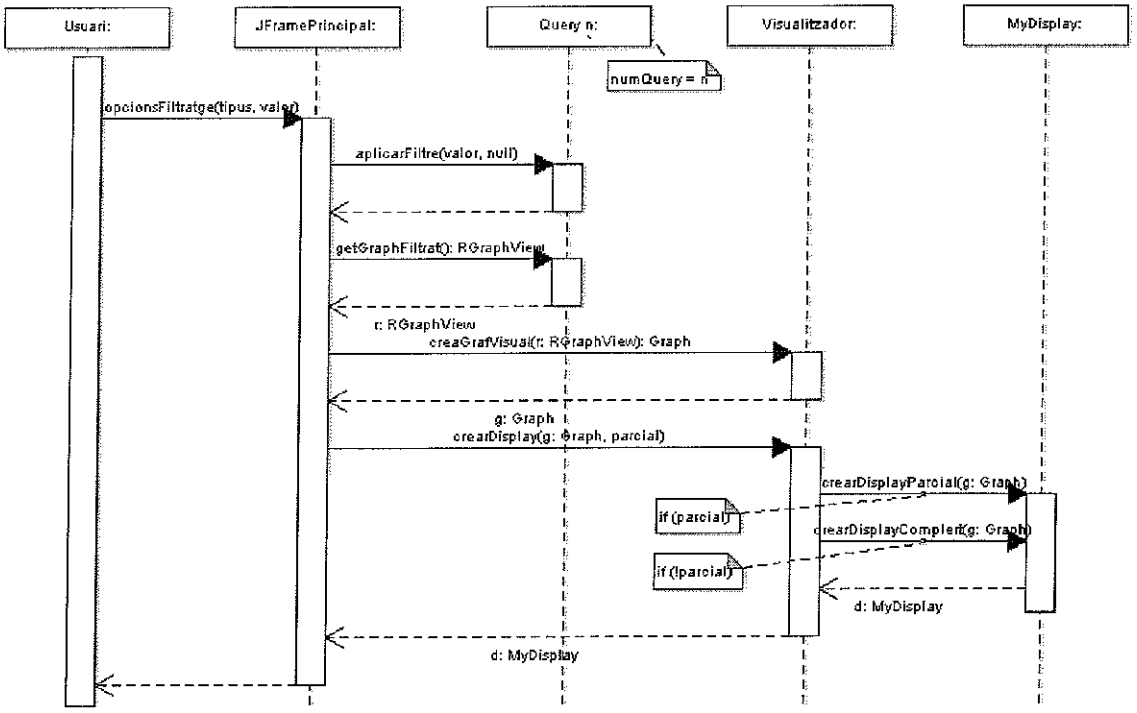


Figura 4.20: Diagrama de seqüència *Opcions filtratge*

Visualitzar informació node/aresta

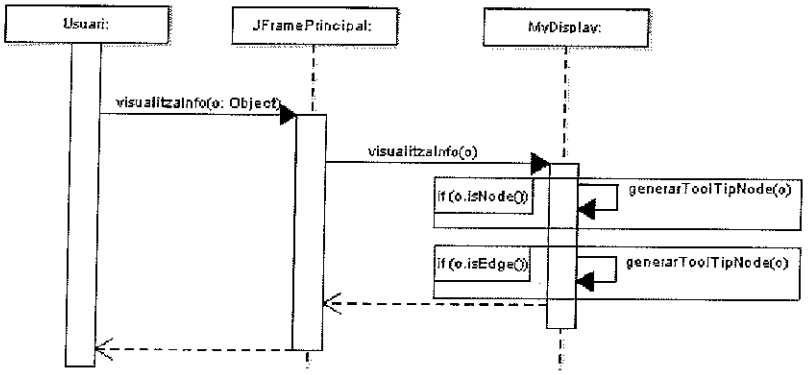


Figura 4.21: Diagrama de seqüència *Visualitzar informació*

Cerca informació

De la mateixa manera que en els diagrames anteriors, els dos casos d'ús que conformen el cas genèric *Cerca informació,* tenen la mateixa seqüència d'operacions i utilitzen les mateixes classes. Com a exemple, es defineix a continuació la seqüència pel cas *Cercar persones.* L'única diferència respecte a l'altre cas d'ús, es troba en les operacions *getPathListPers()* i *visualitzaLlistaPersones().* La distinció tan sols es produeix en la nomenclatura ja que l'objectiu és el mateix.

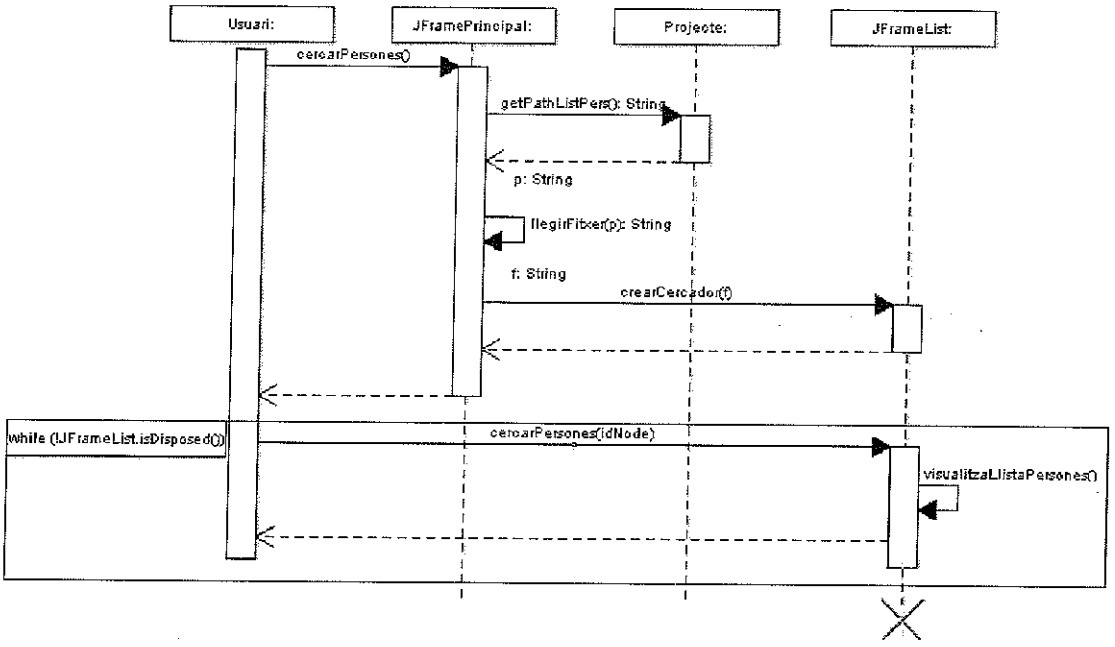


Figura 4.22: Diagrama de seqüència *Cercar persones*

Capítol 5

Implementació

Seguint amb el cicle de vida clàssic d'un sistema software, i després d'haver definit detalladament les tres primeres etapes d'anàlisi, especificació i disseny; és necessari prosseguir amb aquesta etapa. La implementació consisteix en codificar, utilitzant el llenguatge de programació elegit, tot els components de l'aplicació amb l'objectiu d'oferir les funcionalitats requerides. D'aquesta manera, el conjunt del capítol ofereix una descripció de tot allò relacionat amb la implementació de l'aplicació que mereix una especial atenció.

5.1 Introducció

Primer de tot, és necessari definir els recursos de software que han de ser utilitzats en la codificació per tal d'entendre la descripció de la implementació que s'exposarà en els següents apartats.

5.1.1 Llenguatge de programació

Havent elegit un patró de disseny de *Crida i Retorn* basat en orientació a objectes, el llenguatge que més encaixa amb aquest projecte és *Java* de Sun MicroSystems [18] que garanteix la portabilitat atès que utilitza llibreries pròpies independents del sistema operatiu on està corrent l'aplicació. La versió de la màquina virtual que s'ha utilitzat és la 1.6 que conté força millores respecte l'anterior.

D'altra banda, el fet que els components s'estructurin en classes garanteix la canviabilitat. És a dir, s'ofereix la possibilitat de realitzar qualsevol ampliació o modificació en un futur. També n'assegura la utilitat ja que, cadascuna de les classes d'aquesta implementació, poden ser utilitzades en el disseny d'un altre sistema software.

Pel que fa la gestió de memòria en *Java,* implica certa seguretat degut a què no existeix aritmètica de punters, el programador no pot controlar com es gestiona la memòria i, finalment, el rang en els arrays està controlat pel propi llenguatge.

No obstant, cal fer menció d'un inconvenient important en *Java.* El fet que la gestió de memòria tan sols la realitzi el propi llenguatge provoca que, sovint, es produeixin excessos d'utilització de memòria per un mal funcionament del recol·lector de brossa. Cal tenir en compte, en cada moment, com es realitza la implementació per tal d'evitar aquests problemes.

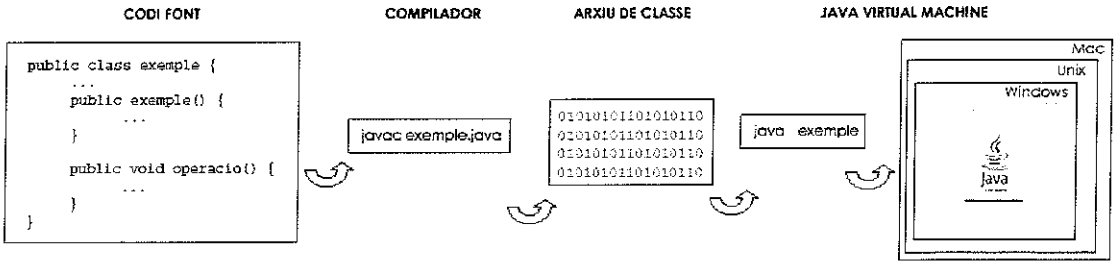


Figura 5.1: Exemple de funcionament de *Java*

5.1.2 Llibreries

Tot i que el llenguatge *Java* utilitza llibreries pròpies per tal de realitzar certes funcions, en aquest projecte, ha estat necessària la utilització de llibreries externes addicionals per tal d'aconseguir el màxim d'objectius establerts per l'aplicació. En total, s'ha incorporat quatre llibreries que són les següents:

- *Jdex* [6]: Llibreria creada pel grup de desenvolupament de DAMA-UPC [17]. Permet la creació de Bases de Dades orientades a graf i la posterior navegabilitat a través d'aquest model per tal d'executar-hi consultes i crear nous grafs resultants.
- *Prefuse* [24]: Llibreria de codi obert que conté un conjunt d'eines per a la creació, manipulació i visualització de grafs.
- *JExcelApi* [21]: Llibreria gratuïta que permet llegir, escriure i modificar arxius de Microsoft Excel dinàmicament.
- *Jakarta Commons Configuration* [15]: Llibreria gratuïta que permet gestionar les configuracions de les aplicacions creades en Java de manera avançada.

5.2 Capa de dades

5.2.1 Introducció

Les classes que componen la capa de dades són *Magatzem* i *NextLineData*. La classe *Magatzem* és l'encarregada de llegir la Base de Dades relacional en format *Comma Separated Values* (CSV) i transformar-la al nou model de dades orientat a graf que ofereix la llibreria *Jdex*.

La classe *NextLineData* és utilitzada per emmagatzemar cada fila del fitxer CSV que representa un diagnòstic o procediment aplicat en una determinada alta hospitalària d'un pacient en concret. Cada individu registrat a la Base de Dades relacional té un conjunt de files que corresponen a una o vàries altes hospitalàries. Cada fila es representa mitjançant *NextLineData* amb l'objectiu d'ordenar-ne els ingressos cronològicament i crear la nova Base de Dades orientada a graf amb les dades correctes.

5.2.2 Base de Dades relacional

Per tal que el model de Base de Dades orientat a graf de *Jdex* pugui ser generat satisfactòriament, cal tenir clar com ha de ser la Base de Dades relacional inicial. A continuació, es presenta quin és el procés de creació i la descripció i estructura dels elements que la formen.

5.2.2.1 Origen de les dades

Si disposem del RHT [25] i de les AH de l'Hospital Universitari de Bellvitge i de l'ICO, aleshores tenim el conjunt de dades idoni per a ser tractat i poder obtenir la informació que permet dur a terme l'avaluació del conjunt de procediments diagnòstics i terapèutics implicats en els ingressos hospitalaris corresponents als diagnòstics de tots els tumors.

El RHT representa el conjunt de tumors diagnosticats. D'altra banda, les AH són la Base de Dades que conté el conjunt d'altes hospitalàries per a cada pacient registrat. Ambdues fonts de dades provenen de l'Hospital Universitari de Bellvitge i de l'ICO.

Cada tumor registrat al RHT conté un identificador de persona que representa l'individu que ha estat diagnosticat amb aquest tumor. Partint d'aquest identificador, es pot dur a terme un encreuament de dades amb les AH del mateix centre sanitari i, consegüentment, obtenir tots aquells ingressos hospitalaris que pertanyen a l'individu i que van ser registrats després d'haver estat diagnosticat amb aquest tumor.

D'aquesta manera, la combinació d'aquestes dues fonts de dades és una Base de Dades relacional que conté la informació del conjunt de tumors diagnosticats, a més de contenir la informació d'aquelles altes hospitalàries associades a aquestes persones encara que no totes tinguin com a causa d'ingrés el tumor diagnosticat.

5.2.2.2 Construcció de la Base de Dades

La combinació del RHT i les AH dóna lloc a una nova Base de Dades relacional d'una magnitud considerable. És per aquest motiu que, el grup que desenvolupa el rol de client, ha cregut oportú realitzar una sèrie de filtratges i transformacions que en redueixin les dimensions i facilitin la tasca de validació de l'eina implementada.

S'ha establert que tan sols es mostrin aquelles persones que només han estat diagnosticades d'un tumor, obviant així les que tenien més d'un tumor diagnosticat al llarg de la seva vida. D'aquesta manera, l'estudi es centra en un localització tumoral i tota la història d'ingressos hospitalaris dels pacients diagnosticats d'aquest tumor que han estat registrats després del diagnòstic. Cal recordar que els diagnòstics de tumors de tots els pacients i les altes hospitalàries que hi estan relacionades pertanyen única i exclusivament a l'Hospital Universitari de Bellvitge i a l'ICO.

Per tant, tan sols ha estat necessari prendre les dades del registre i la Base de Dades esmentats en el punt anterior i filtrar-ne tots aquells pacients que no complien amb els requisits citats. Per a simplificar-ho encara més, s'ha decidit obviar aquelles variables de les AH que no aportaven cap tipus d'ajuda a l'avaluació dels procediments implicats per a l'extracció de regles i algorismes pel software ASEDAT [2].

5.2.2.3 Descripció de la Base de Dades

Després d'haver especificat i dissenyat la Base de Dades ha estat necessari construir-la de la manera que s'ha descrit en aquesta secció. Actualment, es disposa d'una Base de Dades formada per 11 taules relacionals diferents que, a continuació, se'n presenta la descripció i les variables que contenen:

Taula *serveis*

Descripció **Conté el conjunt de serveis sanitaris que poden ser utilitzant durant l'aplicació de procediments a pacients i els centres on s'imparteixen aquests serveis.**

Variables	Descripció	Tipus	Exemple
idServei	Identificador del servei	String	ACV
idCentre	Identificador del centre associat al servei	Integer	11
servei	Servei que pot ser utilitzat durant l'aplicació de procediments a pacients	String	ANG.C. VASCULAR

Taula **circumsAdmi**

Descripció **Conté el conjunt de circumstàncies d'admissió d'un ingrés**

Variables	Descripció	Tipus	Exemple
idCAd	Identificador de la circumstància d'admissió	Integer	1
circumAd	Circumstància d'admissió d'un pacient en un ingrés	String	Programat

Taula ***circumsAlta***

Descripció **Conté el conjunt de circumstàncies per les quals un pacient rep l'alta hospitalària.**

Variables	Descripció	Tipus	Exemple
idCAl	Identificador de la circumstància d'alta	Integer	1
circumAl	Circumstància d'alta d'un pacient	String	Domicili

Taula ***tipusIngres***

Descripció **Conté el conjunt de diferents tipus d'ingrés que poden existir.**

Variables	Descripció	Tipus	Exemple
idTI	Identificador del centre hospitalari	Integer	1
tipusIngres	Tipus d'ingrés	String	Cirurgia general

Taula **procedencia**

Descripció **Conté el conjunt de possibles llocs de procedència del pacient ingressat.**

Variables	Descripció	Tipus	Exemple
idProc	Identificador del servei	Integer	1
procedencia	Identificador del centre associat al servei	String	Propi hospital

Taula **altesHospitalaries**

Descripció **Conté el conjunt de les altes hospitalàries d'entre el total de tots els individus.**

Variables	Descripció	Tipus	Exemple
idLinia	Identificador de fila de la taula	Integer	8
idPers	Identificador de persona ingressada	Integer	123654
numIngres	Identificador de l'ingrés de la persona ingressada	Integer	53882
idServei	Identificador del servei utilitzat durant l'ingrés	String	BRA
idProc	Identificador del lloc de procedència del pacient ingressat	Integer	1
centre	Centre on ha estat ingressat el pacient	Integer	
idCAd	Identificador de la circumstància d'admissió del pacient	Integer	1
idTI	Identificador del tipus d'ingrés	Integer	3
dataIngres	Data en què ha estat ingressat el pacient	String	22/9/2003 0:00:00
idEA	Identificador de l'estat d'alta	Integer	3
idCAl	Identificador de la circumstància d'alta	Integer	1
dataAlta	Data en que s'ha donat d'alta el pacient	String	23/9/2003 0:00:00
codiICD9	Diagnòstic o procediment aplicat al pacient en aquest ingrés	String	174.8
tipus	Defineix si el codi ICD9 és un diagnòstic o un procediment, a més a més, d'especificar si es principal o secundari	String	P
idTipusCirurgia	Identificador del tipus de cirurgia	Integer	3

Taula **estatAlta**

Descripció **Conté el conjunt de possibles estats en què es troba el pacient quan rep l'alta hospitalària.**

Variables	Descripció	Tipus	Exemple
idEA	Identificador de l'estat d'alta	Integer	1
estatAlta	Estat en què es troba el pacient en el moment de donar-li l'alta hospitalària	String	Curació

Taula *centres*

Descripció **Conté el conjunt de centres sanitaris i les seves àrees.**

Variables	Descripció	Tipus	Exemple
idCentre	Identificador del centre sanitari	Integer	11
idAH	Identificador de l'àrea	String	H08000924
idUP	Identificador de la unitat proveïdora	Integer	148
centre	Centre sanitari	String	HOSPITAL BELLVITGE

Taula **icd9**

Descripció **Conté el total de codis i descripcions de cada diagnòstic i procediment en l'estàndard de codificació ICD9.**

Variables	Descripció	Tipus	Exemple
codi	Codi que segueix l'estàndard ICD9 utilitzat per a identificar malalties	String	001
tipus	Defineix si la malaltia és un diagnòstic o un procediment	String	D
descrLlarga	Descripció llarga de la malaltia	String	COLERA PER VIBRIO CHOLERAE
descrMitja	Descripció mitja de la malaltia	String	COLERA D/A VIB CHOLERAE
descrCurta	Descripció curta de la malaltia	String	COLERA D/A VIB

Taula **areesHospitals**

Descripció **Conté el conjunt de les àrees hospitalàries.**

Variables	Descripció	Tipus	Exemple
idAH	Identificador de l'àrea hospitalària	String	H25001678
areaHospital	Àrea hospitalària	String	CLINICA DE REPOS S. JOSEP ORIO

Taula **tumorsIncidents**

Descripció **Conté el conjunt d'individus diagnosticats d'un tipus de tumor.**

Variables	Descripció	Tipus	Exemple
idPers	Identificador de la persona amb diagnòstic d'un tumor	Integer	223555
idLinia	Identificador de la fila de la taula	Integer	2
diaIncidencia	Dia que es diagnostica el tumor en el pacient	Integer	12
mesIncidencia	Mes que es diagnostica el tumor en el pacient	Integer	7
anyIncidencia	Any que es diagnostica el tumor en el pacient	Integer	2005
idMetodeDX	Identificador del mètode de diagnòstic	Integer	7
codiICD9	Codi del tumor diagnosticat en el pacient codificat amb l'estàndard ICD9	Integer	182
especificacioTumor	Detall del tipus de tumor que ha estat diagnosticat	Integer	0
histologia	Histologia del tumor	Integer	8380
indexComportament	Index de comportament del tumor	Integer	3
indexCharlson	Index de comorbiditat de Charlson	Integer	
indexElixhauser	Index de comorbiditat de Elixhauser	Integer	

5.2.2.4 Format de la Base de Dades

Resulta molt interessant el fet de disposar de totes les dades referents als ingressos hospitalaris de cada centre sanitari de Catalunya unides en un sol registre. Aquesta és la idea del CMBD [14] que requereix que totes les AH de qualsevol centre sanitari segueixin l'especificació marcada per aquest registre.

Atès que la nova Base de Dades relacional creada en aquest projecte té el seu origen en la combinació del RHT [25] i les AH, es pot afirmar que també compleix amb el què indica l'especificació esmentada. Així doncs, l'aplicació desenvolupada està preparada per a ser utilitzada partint de qualsevol conjunt de dades que tingui el seu origen en alguna o totes les AH del total de centres sanitaris del territori català.

Pel què fa la implementació de la nova Base de Dades relacional, podem dir que conté un total de 11 fitxers que segueixen el format *Comma Separated Value* (CSV). Aquests, corresponen a les 11 taules relacionals de la qual està formada la Base de Dades. Cadascun d'aquests arxius, ha de tenir un nom específic per tal què l'aplicació pugui localitzar-lo i carregar-ne les dades corresponents. En la següent taula, es pot veure quin és el nom que ha de tenir cadascun d'aquests fitxers juntament amb la taula relacional de la Base de Dades que li correspon.

Taula relacional	Nom del fitxer CSV
tipusIngres	tipus_ingres.csv
serveis	serveis.csv
centres	centres.csv
areesHospitals	arees_hospitals.csv
estatAlta	estat_alta.csv
circumsAdmi	circuns_admin.csv
circumsAlta	circums_alta.csv
icd9	icd9.csv
procedencia	procedencia.csv
altesHospitalaries	AH_DEX.csv
tumorsIncidents	TUMORS_DEX.csv

La llibreria *Jdex* ofereix la possibilitat de carregar les dades registrades en una Base de Dades relacional emmagatzemada en arxius CSV, de manera que tan sols és necessari utilitzar aquesta funcionalitat per obtenir les dades d'aquests fitxers i prosseguir amb la construcció de la nova Base de Dades orientada a graf.

5.2.2.5 Observacions generals

Cal tenir present que poden existir alguns errors en els resultats que s'obtenen mitjançant l'aplicació implementada. L'origen rau en el fet que, la Base de Dades relacional construïda, conté tan sols les altes hospitalàries registrades en els centres sanitaris de l'Hospital Universitari de Bellvitge i de l'ICO. Encara que el conjunt d'ingressos hospitalaris posteriors al diagnòstic d'un tumor en un pacient és real, es pot ocasionar el fet que una persona diagnosticada d'un tumor hagi estat ingressada en altres centres sanitaris que no siguin els esmentats i, per tant, no es disposi de totes les altes hospitalàries posteriors al diagnòstic d'aquest tumor. D'altra banda, també és possible que un pacient hagi estat diagnosticat d'un tumor en un centre sanitari aliè als citats anteriorment i, en canvi, els ingressos hospitalaris posteriors a aquest diagnòstic es registrin a l'Hospital Universitari de Bellvitge o a l'ICO. En qualsevol dels dos casos, no es disposarà de totes les dades necessàries per a realitzar un estudi complert.

De totes maneres, aquest esdeveniment no és gaire freqüent per la qual cosa no es pot considerar un problema en l'anàlisi de resultats sinó un punt a tenir en compte. A més, tal i com està previst, en un futur s'utilitzarà el sistema software partint d'una Base de Dades que agruparà totes les AH de tots els centres sanitaris de Catalunya. Aleshores, s'haurà solucionat aquest problema ja que es disposarà de tots els ingressos hospitalaris d'un pacient encara que hagi estat ingressat en diversos centres del territori català.

També cal esmentar el fet que els identificadors de persona no són reals. Tot i així, permeten realitzar la funció d'identificador que és el que es necessita en aquest projecte. És a dir, enlloc de tenir el número de seguretat social o el número del document d'identitat, es disposa d'un identificador fictici amb el qual, aquells qui tenen permís, poden obtenir l'identificador real i consultar les dades personals necessàries.

En aquest cas, la projectista no ha tingut accés a les dades personals degut a la llei orgànica de protecció de dades (LOPD) que regeix des del 1999.

5.2.3 Base de dades orientada a graf

Un cop s'ha definit quin ha de ser el format inicial de la Base de Dades relacional és el moment de descriure el procés de transformació cap a la nova Base de Dades orientada a graf mitjançant la llibreria *Jdex [6].* El motiu pel qual s'ha escollit aquesta llibreria per a construir una nova Base de Dades que segueixi el model orientat a graf és el gran volum de dades a tractar. *Jdex* permet carregar tota una Base de Dades relacional dins una única estructura en forma de graf on els nodes equivalen a una taula i les arestes a les seves relacions. A més, permet realitzar consultes sobre aquest nou model de dades obtenint com a resultat un nou graf i, alhora, dóna la possibilitat d'executar consultes sobre una gran quantitat de dades de manera eficient ja que en un model relacional seria pràcticament impossible.

En el desenvolupament d'aquesta aplicació, la classe *Magatzem* és l'encarregada de llegir totes les taules de la Base de Dades relacional inicial i transformar-la en el nou model de dades orientat a graf.

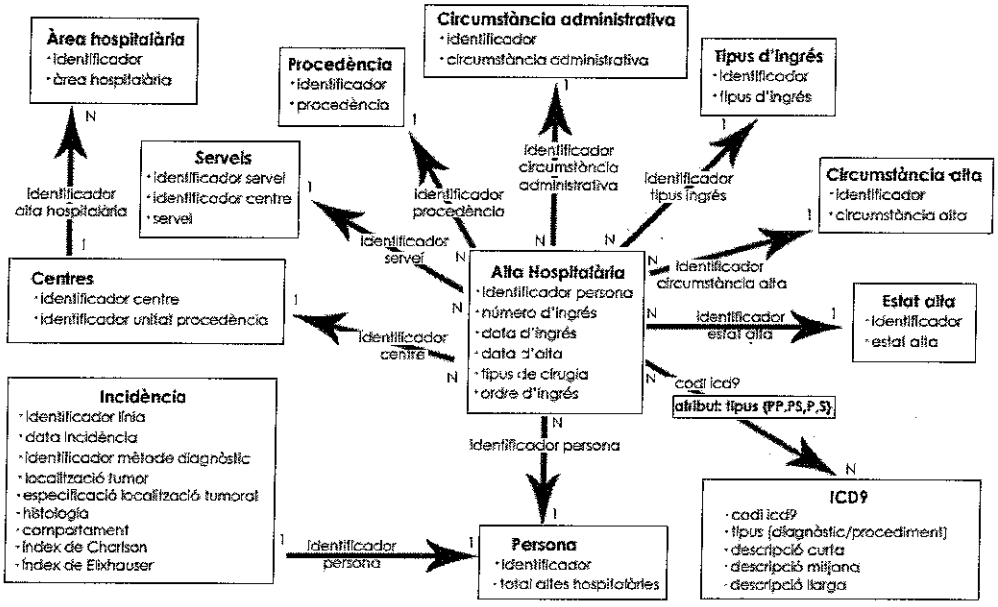


Figura 5.2: Model de dades orientat a graf del sistema software desenvolupat

A la Figura 5.2, podem veure com cada quadre de la imatge que representa una entitat equival a un node i, les fletxes, són les arestes que representen les relacions entre els diferents nodes. Dins de cada aresta, es pot veure per quin atribut es relacionen les dos entitats del seu extrem. Per a cada parell d'entitats que en el model relacional estaven relacionades mitjançant una clau foran que era un identificador, es crea una aresta entre elles dues i, consegüentment, s'eliminen l'identificador esmentat. Les pròpies arestes poden contenir atributs amb l'objectiu d'assignar informació a la relació entre dos nodes.

5.2.4 Càrrega de la Base de Dades orientada a graf

Finalment, després d'haver definit els dos models de dades i la corresponent construcció de la Base de Dades orientada a graf, cal fer menció de com es carrega el model de dades orientat a graf en l'execució de l'aplicació final. El sistema software desenvolupat és l'encarregat de comprovar l'existència de la Base de Dades orientada a graf.

Per una banda, si es dóna el cas que encara no ha estat creada, caldrà dur a terme la lectura de la Base de Dades relacional en format CSV i, acte seguit, crear el nou model de dades orientat a graf. Un cop

hagi estat creada aquesta segona Base de Dades, l'aplicació n'emmagatzema una imatge a disc. Finalment, l'últim pas consisteix en carregar la nova imatge creada per tal d'accedir a qualsevol de les seves dades durant l'execució de l'aplicació.

D'altra banda, si ja existeix una imatge a disc de la Base de Dades orientada a graf, el sistema software es limitarà a carregar-la per a permetre'n l'accés a les seves dades durant tota l'execució del programa.

5.2.5 Fitxers de configuració

Degut a què l'aplicació treballa en un entorn de projectes que en permet la seva creació i gestió, és necessari utilitzar alguna metodologia per a desar la informació associada a la configuració de cada projecte.

És per aquest motiu que s'utilitza la llibreria *Jakarta Commons Configuration [15]*, que permet carregar i emmagatzemar a disc fitxers de configuració en format XML. La llibreria permet tant l'accés a aquest tipus d'arxius com la seva escriptura si és pertinent. En el sistema software implementat, s'utilitzen dos fitxers de configuració inicials. A continuació, es pot observar la descripció de cadascun d'aquests arxius:

Global.xml

Fitxer de configuració de l'aplicació. Conté els directoris on s'emmagatzemen la Base de Dades relacional, la Base de Dades orientada a graf i els projectes. En un futur, pot permetre emmagatzemar qualsevol altra informació necessària per a la correcta configuració de l'aplicació. El seu format és el següent:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<Global>
    <path>
        <pathDEX>dex\</pathDEX>
        <imatgeDEX>dex\data.dex</imatgeDEX>
        <modelCSV>csv\</modelCSV>
        <projects>projects\</projects>
    </path>
</Global>
```

Projecte.xml

Fitxer de configuració dels projectes en general. Cada cop que s'emmagatzema un projecte a disc, es copia aquest arxiu al directori del projecte i se li assigna la informació necessària. Així doncs, aquest fitxer XML conté els directoris dels fitxers generats en les consultes a petició de l'usuari, a més del directori on es desarà el fitxer de configuració del projecte. El seu format és el següent:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<Projecte>
    <tumor> </tumor> <!-- Localització tumoral que cal assignar al projecte -->
    <path>
        <config>config/confProj.xml</config>
        <graphml>graphml</graphml>
        <report>report</report>
        <listPers>listPers</listPers>
        <listIngr>listIngr</listIngr>
        <excel>excel</excel>
    </path>
</Projecte>
```

5.3 Capa de domini

En el sistema software implementat, la capa de domini engloba totes aquelles classes que permeten l'execució de les cinc consultes contra la Base de Dades orientada a graf, a més de totes aquelles que contenen operacions útils per a la correcta realització de les funcions d'aquesta capa. És doncs, una de les parts més importants en la codificació del conjunt del programa.

5.3.1 Navegació per la Base de Dades orientada a graf

Cal tenir present que la Base de Dades orientada a graf registra una gran quantitat de dades. És per aquest motiu que, sovint, no és necessari prendre-les totes per a obtenir informació útil en cada consulta. De fet, existeixen vàries variables que no seran utilitzades en la construcció dels grafs resultants de les consultes ja que l'objectiu en aquest sistema software és obtenir informació per a l'avaluació dels procediments aplicats. Com és d'esperar, les dades que més vegades seran visitades són les que fan referència a l'entitat que representa les altes hospitalàries i, dins d'aquesta, les variables que registren els procediments, diagnòstics i persones associades.

No obstant, en un futur poden ser implementades noves consultes on l'objectiu sigui un altre, i per tant, sigui necessari l'accés a altres variables emmagatzemades, de manera que s'ha decidit no obviar cap dada en el moment de carregar la Base de Dades orientada a graf.

Així doncs, la navegació per tal de construir el nou graf resultant que es realitza en cada consulta comença cercant les incidències del tumor indicat pel projecte obert, passant per la persona que li corresponen a cada incidència, les altes hospitalàries posteriors a la data d'incidència del tumor i tots els procediments i diagnòstics associats.

En el gràfic de la Figura 5.3, es pot veure quina és la navegació que es produeix a través de les diferents entitats de la Base de Dades orientada a graf en cada execució de les diferents consultes de l'aplicació. Tal i com s'ha dit abans, existeixen moltes variables que no són accedides pels motius exposats.

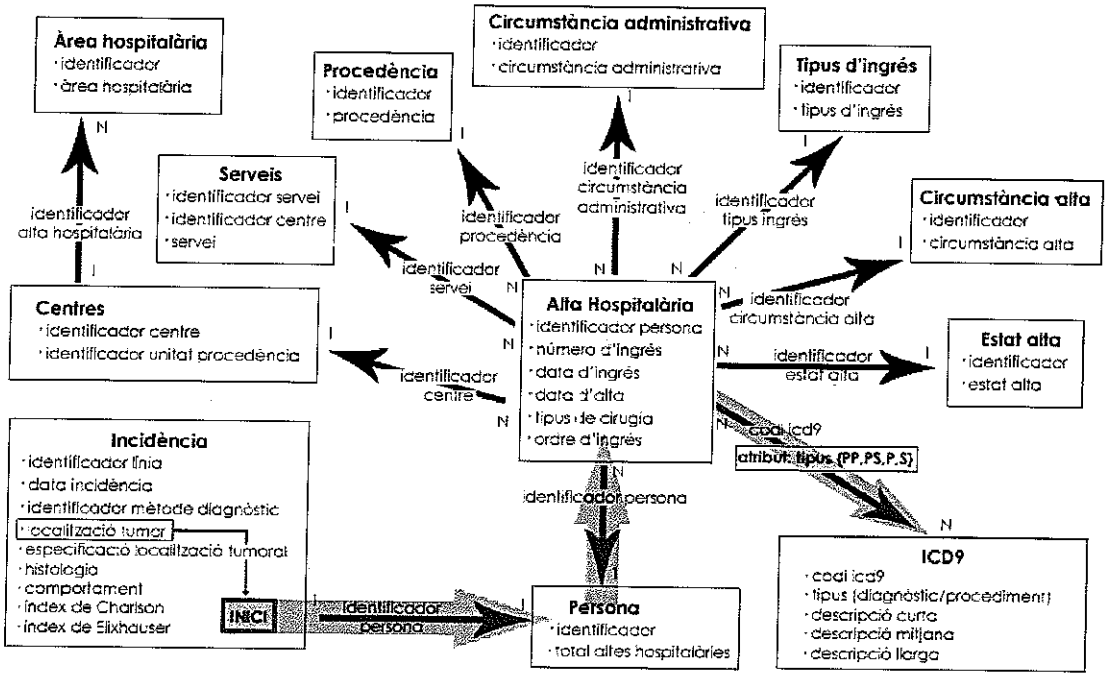


Figura 5.3: Navegació per la Base de Dades orientada a graf

Finalment, la Figura 5.4, mostra un exemple del què seria el conjunt de nodes i arestes associades al diagnòstic d'un tumor en concret d'una persona. Cada persona diagnosticada d'un tipus de tumor determinat es representa mitjançant una instància de l'entitat *Incidència*. Gràcies a aquest exemple, es pot observar el camí que segueix la navegació per la Base de Dades orientada a graf en una de les moltes iteracions per recórrer el total de diagnòstics d'un tipus de tumor determinat.

En l'exemple, es suposa que aquesta localització tumoral és *152 (càncer d'intestí prim)* i es pren una incidència qualsevol d'aquest tumor. La navegació correspon a una iteració sobre el conjunt de totes les incidències d'un tumor en concret, en aquest cas, el càncer d'intestí prim. Es pot veure com, per cada alta hospitalària associada a la persona amb l'identificador *902551* diagnosticada del tumor *152*, s'obtenen tots els procediments i diagnòstics que té registrats. Les arestes que uneixen les altes hospitalàries i els procediments i diagnòstics representats mitjançant l'entitat *ICD9*, contenen un atribut que indica si es tracta d'un procediment principal (PP), un procediment secundari (PS), un diagnòstic principal (P) o un diagnòstic secundari (S).

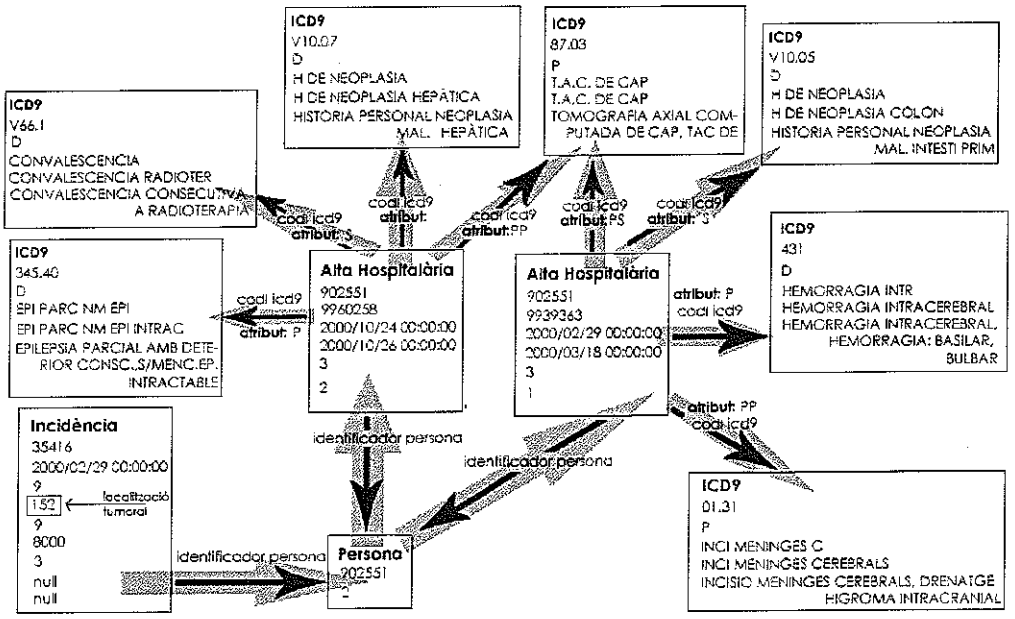


Figura 5.4: Exemple de navegació partint d'una incidència tumoral *152*

5.3.2 Construcció del graf resultant

En cadascuna de les cinc consultes implementades existeixen tres operacions per tal de construir el graf que conté totes les dades i escriure'n els resultats en fitxers addicionals. Aquestes tres funcions són la base de la construcció del graf resultant i, és per aquest motiu que, a continuació se les descriu amb detall.

5.3.2.1 Operació per a la construcció del graf: *creaGraf()*

En aquesta operació, es construeixen els nodes i les arestes del graf resultant de l'execució de qualsevol de les cinc consultes. A més, s'introdueixen les dades pertinents dins de cada node o aresta del graf.

Per tal de poder construir el graf correctament, és necessari realitzar una iteració per totes les instàncies de la classe *Incidència* que tenen una localització tumoral que coincideix amb la indicada pel projecte obert.

En cadascuna d'aquestes iteracions, la instància de *Incidència* és l'arrel del graf de la Base de Dades que cal llegir en pre-ordre, de la mateixa manera que s'ha mostrat en l'apartat de navegació. Per cada node o aresta que es llegeixi i segons la informació que es desitgi representar en el nou graf resultant, caldrà crear un node o aresta d'un tipus o d'un altre.

Tal i com ja s'ha vist a la secció 3.2 a la pàgina 24 que fa referència a la descripció de les diferents consultes, una de les informacions que es mostren dins de les arestes és la quantitat de persones agrupades en el node del seu extrem i el percentatge que representa respecte els seus veïns. És evident que fins que no s'ha recorregut tot el graf de la Base de Dades, no es pot conèixer quantes vegades es repeteix un diagnòstic, un procediment o qualsevol altra informació que necessita llegir tot el graf per a ser obtinguda.

D'aquesta manera, en aquesta operació, tan sols es pot construir el graf resultant amb les dades que es puguin extreure directament de la Base de Dades orientada a graf. Aquelles que requereixin un mínim de càlculs hauran de ser introduïdes en les següents operacions.

5.3.2.2 Operació per a completar el graf: *completaPercentatges()*

Un cop s'ha recorregut tota la Base de Dades orientada a graf ja es coneixen les dades necessàries per a calcular aquella informació que mancava per introduir en el graf resultant de l'operació anterior. És el moment de recórrer el graf resultant en pre-ordre per tal de llegir les dades introduïdes en el pas anterior i assignar els percentatges dels nodes i arestes que completaran la informació total d'aquest graf.

Paral·lelament a aquesta acció, es crea una nova estructura de graf que permet ordenar els nodes del mateix nivell segons el percentatge que representen. Aquest nou graf conté, a cada node, un atribut de la classe *ProcDiag* que disposa d'un identificador de node del graf resultant, del percentatge que representa aquest node entre els nodes del seu mateix nivell, i finalment, d'un atribut de la classe *ArrayList*. Cal afegir que, aquest nou graf, té la mateixa forma que el graf resultant creat en l'anterior operació ja que la construcció d'aquest segon es fa al mateix temps que es llegeix el primer. Cada node representat per la classe *ProcDiag* conté un identificador, un percentatge i un apuntador a una *ArrayList*. Cada *ArrayList* conté un conjunt d'instàncies de la classe *ProcDiag* que representen els fills del node que conté aquesta llista. D'aquesta manera, és possible ordenar tots els nodes d'un mateix nivell segons el seu percentatge. A la Figura 5.5, s'hi observa l'estructura d'aquest segon graf utilitzat per ordenar.

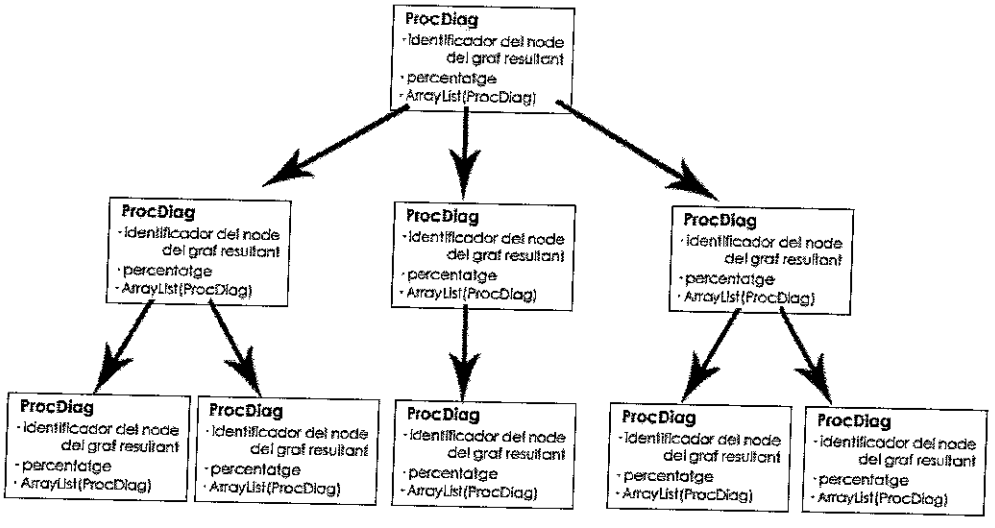


Figura 5.5: Graf auxiliar per tal d'ordenar els nodes del mateix nivell

Tal i com podem veure, per tal d'ordenar els nodes d'un mateix nivell segons el seu percentatge tan sols és necessari ordenar cada *ArrayList* de cada node amb el mètode *sort*, el qual s'encarrega d'ordenar tots els atributs de classe *ProcDiag* que conté.

Finalment, cal dir que la llibreria *Jdex [6]* no ofereix encara cap mètode que permeti ordenar els nodes d'un mateix nivell segons un dels seus atributs, per la qual cosa no ha estat possible realitzar-ho d'una manera més còmoda que l'exposada.

Una altra solució per la qual es podria haver optat és la de tornar a generar un graf mitjançant *Jdex*, de la mateixa manera que s'ha creat el graf resultant de l'execució de la consulta, amb l'objectiu d'introduir-hi els elements degudament ordenats. No obstant, aquesta llibreria no insereix els nodes un darrere de l'altre, per la qual cosa no és possible realitzar insercions ordenades.

5.3.2.3 Operació per a escriure als fitxers generats: *escriureResultat()*

Un cop s'ha construït el graf resultant i s'ha generat una nova estructura de graf que conté els nodes ordenats, és el moment d'escriure els resultats en aquells fitxers addicionals que l'usuari hagi sol·licitat. És aquí on s'utilitza el graf auxiliar generat en l'operació anterior. Per tal de mostrar els resultats ordenats segons el percentatge que representen, es fa una lectura d'aquest graf auxiliar en pre-ordre i, cada cop que es llegeix un node de classe *ProcDiag*, es pren l'identificador de node que conté i es realitza la cerca en el graf resultant generat en la primera operació. Un cop s'ha pogut accedir a les dades contingudes en aquest node, es prenen aquelles que siguin necessàries per a la construcció de cada fitxer i s'hi escriu la informació respectiva a cadascun d'ells de manera ordenada.

5.3.3 Cost de la construcció del graf resultant

Tal i com s'acaba d'explicar, per tal de construir el graf resultant, són necessàries com a mínim dues operacions. La primera per a crear-ne el graf i, la segona, per a completar-ne la informació. En el pitjor dels casos però, l'usuari desitjarà obtenir resultats en altres formats que no siguin tan sols la visualització del graf resultant i, per tant, caldrà generar nous fitxers.

Considerant n el total dels nodes de la Base de Dades orientada a graf, en el pitjor dels casos el cost serà:

$$\text{cost } \textit{creaGraf()} + \text{cost } \textit{completaPercentatges()} + \text{cost } \textit{escriureResultat()}$$

$$O(n) + O(n) + O(n) = O(n^3)$$

Així doncs, el cost total és de tipus cúbic. S'interpreta que en el cas pitjor, si es dobla la mida de la Base de Dades, aleshores el cost es veu multiplicat per 8, i per tant, el temps d'execució també es veu augmentat de manera similar.

Cal dir que un cost polinòmic no és dolent, tot i què influeix força en el temps final d'execució. No obstant, gràcies a la llibreria *Jdex [6]*, aquesta operació es realitza en un temps força ràpid si s'executa l'aplicació en una màquina que compleixi els requisits especificats. De totes maneres, tal com s'ha explicat anteriorment, no existeix una manera més eficient de fer-ho ja que per a completar els resultats és necessari haver llegir tota la Base de Dades orientada a graf dues vegades.

5.3.4 Definició de l'exportació del graf resultant

Finalment, un cop s'ha construït el graf que sorgeix del resultat d'una consulta determinada, és necessari definir com es mostrarà i quina estructura seguirà la informació extreta del graf resultant en la futura visualització del graf. La classe que s'encarrega de realitzar aquesta funció és la que rep el nom de *MyExport*. Aquesta, implementa la classe *Export* de la llibreria *Jdex [6]* i permet la definició de les propietats dels nodes i les arestes segons el tipus al qual pertanyin. Aquestes propietats són el color, la forma, la mida, el tipus de text i l'estructura de la informació que s'hi visualitzarà.

Les operacions implementades en aquesta classe i la funció que desenvolupen es mostren a continuació:

prepare(View v): obté tots els tipus de nodes i arestes que es poden trobar en un graf resultant de la llibreria *Jdex*.

getNode(Node n): obté un node del graf resultant i li assigna totes les propietats pertinents segons el tipus al qual pertanyi, a més a més, d'assignar-hi el text a visualitzar amb l'estructura que li correspongui.

getEdge(Edge e): realitza la mateixa funció que l'anterior operació però per a arestes

Tan sols cal crear una instància d'aquesta classe quan s'hagi creat el graf resultant. Més endavant, quan es desitgi visualitzar el graf, serà necessari utilitzar la instància de *MyExport* amb l'objectiu d'executar-ne les seves operacions i definir com serà la visualització del graf. Les operacions *getNode* i *getEdge* s'executaran una vegada per a cada node o aresta respectivament.

5.4 Capa de presentació

Com en qualsevol implementació basada en una arquitectura en tres capes existeix una última capa utilitzada per a garantir la comunicació entre l'usuari i l'aplicació. Per tant, es pot considerar com un dels pilars de l'aplicació ja que si no es codifica correctament, el sistema software pot resultar totalment inservible.

No obstant, en aquesta part de la implementació, existeixen una gran varietat de classes que implementen cadascuna de les finestres de l'aplicació. Aquestes doncs, no seran descrites en detall atesa l'existència de classes molt més interessants com són totes aquelles que fan referència a la visualització del graf resultant.

5.4.1 Exportació del graf resultant

Cal recordar que, a la capa de domini, s'ha descrit com és el procés de definició de les propietats de l'exportació del graf resultant. A partir d'aquí, la capa de presentació s'encarrega de prendre tot allò que s'ha establert a la classe *MyExport* i ho utilitza amb l'objectiu d'exportar el graf resultant a un tipus de graf que serà la base de la futura visualització. En aquest sistema software, l'usuari pot escollir si desitja que el graf resultant sigui exportat a graf en format XML. En canvi, l'exportació a graf de *Prefuse [24]* és implícita ja que és necessària per tal de visualitzar-ne els resultats en forma de graf.

5.4.1.1 Format XML

Tot i què el graf resultant exportat és XML, cal concretar més i parlar d'un format més específic anomenat *GraphML*. Aquest, permet crear fitxers amb l'objectiu de representar-hi grafs. Permet descriure quina ha de ser l'estructura del graf en una futura visualització i les seves propietats, a més, d'indicar la informació que ha de contenir cada element.

Un cop es disposa d'una instància de la classe *MyExport,* és possible l'exportació del graf resultant a un nou graf amb aquest nou format. Aquesta transformació es realitza gràcies a les operacions de la llibreria *Jdex [6].* D'aquesta manera, tan sols és necessari realitzar una crida a un mètode de *Jdex* passant-li el graf resultant creat amb aquesta mateixa llibreria i la instància de la classe *MyExport*:

```
rgraphvParcial.export(graph, Export.Type.YGRAPHML, extra)
```

on `graph` és el fitxer en format *GraphML* on s'hi escriurà, `rgraphvParcial` és el graf resultant i `extra` la instància de la classe *MyExport.*

Aquesta crida doncs, executa les operacions de la instància `extra`, que cal recordar que eren les `getNode(Node n)` i `getEdge(Edge e)`. Aquestes eren utilitzades per a obtenir la informació a visualitzar i la definició de les propietats de cada node o aresta. Un cop executades les operacions per a tots els nodes i arestes, s'escriu el resultat de l'exportació al fitxer `graph` seguint el format *GraphML.*

5.4.1.2 Graf de *Prefuse*

A diferència del graf en format XML, en aquesta exportació del graf resultant no es crea cap fitxer sinó que tan sols es duen a terme un seqüència d'operacions, mitjançant la llibreria *Prefuse,* que permeten crear la visualització del nou graf.

Així doncs, la classe *Visualitzador* s'encarrega de transformar el graf resultant a un nou graf de *Prefuse.* Aquesta classe tan sols necessita la instància de la classe *MyExport,* que definia com havia de ser l'exportació del graf, i el graf resultant obtingut en l'execució de la consulta. A continuació es pot observar quin és el procés necessari per obtenir la visualització del graf resultant:

1. Es pren la instància de *MyExport* i s'executa l'operació ***prepare(View v)*** per obtenir el conjunt dels tipus diferents de nodes i aresta que poden aparèixer en el graf resultant creat amb la llibreria *Jdex.*

2. Es crea la instància que representa el graf de *Prefuse.* Un graf d'aquesta llibreria està format per una taula on s'hi emmagatzemen totes les dades necessàries per a ser visualitzat posteriorment. Cada fila representa un node o aresta i cada columna representa una variable del node o aresta. També es defineix l'estructura que tindrà.

3. S'introdueix la informació dels nodes i les arestes en el nou graf. Per a fer-ho, és necessari realitzar recórrer tot el conjunt de nodes del graf resultant creat amb *Jdex.* Així doncs, per a cada node caldrà executar l'operació ***getNode(Node n)*** per tal d'obtenir l'estructura i propietats que ha de tenir el node del nou graf de *Prefuse.* De la mateixa manera que es fa pels nodes, també se'n fa per les arestes seguint el mateix procés però, aquest cop, executant l'operació ***getEdge(Edge n)***. A continuació, s'ofereix un exemple d'una fila de la taula del nou graf que representa un node.

Label	Shape	Fill	Stroke	FontColor	Highlight	Width	Height	Id	Directed	Fit	FontWidth
V50: ALT [1254]	SQUARE	FF4500	FF4500	000000	BLUE	2	10	1	TRUE	FALSE	12

Finalment, aquest nou graf ja conté tota la informació necessària i queda preparat per a dur a terme l'últim pas que en genera la visualització.

5.4.2 Visualització del graf exportat

Després d'executar la consulta, obtenir-ne els resultats en forma de graf i exportar-lo a un tipus de graf que en permeti la visualització, ha arribat el moment de descriure'n la implementació de la visualització. Cal remarcar la importància d'aquesta fase per la seva dificultat. No és fàcil tractar amb un gran volum de dades, tenir clar quina informació ha de ser mostrada i quina cal obviar, i finalment, com visualitzar-ho per tal de què sigui entenedor, amigable i, sobretot, útil per una correcta anàlisi dels resultats.

5.4.2.1 Visualització del graf en format XML

El fitxer *GraphML* que s'obté en l'exportació descrita anteriorment és un graf que pot ser visualitzat amb qualsevol aplicació que suporti aquest tipus de fitxers. Un exemple de programa que ho permet és *yEd Graph Editor* [26] desenvolupat amb el llenguatge *Java* [18] per l'empresa *yWorks*.

5.4.2.2 Visualització del graf de *Prefuse*

Un cop disposem del graf construït gràcies al suport de la llibreria *Prefuse*, ja està tot a punt per a visualitzar-lo utilitzant la mateixa llibreria. Concretament, s'ha implementat una classe anomenada *MyDisplay* que hereta els mètodes de la classe *Display* de *Prefuse*, la qual s'encarrega de preparar la visualització i mostrar-la en l'aplicació final.

Així doncs, els passos a seguir per a dibuixar el graf són els següents:

1. Crear una nova instància de la classe *MyDisplay*. Amb això, aconseguim que aquesta classe disposi del graf de Prefuse i d'un booleà que li indica si ha de visualitzar el graf complert, o tan sols, un resum d'aquest.

2. Crear la visualització del graf mitjançant una operació que, segons el valor d'un booleà, mostrarà el graf complert o tan sols un resum d'aquest. Si el booleà indica que s'ha de mostrar el graf resum, aleshores s'executarà l'operació que permet agrupar els nodes i resumir-ne la visualització, a més a més, de permetre'n l'explosió d'aquella part del graf que l'usuari seleccioni. Si indica que cal mostrar-lo sencer, es prosseguirà a construir-ne la visualització de tot el graf resultant. Els passos a seguir en les dues operacions mencionades són els següents:

   (a) Crear i afegir un *renderer* per als nodes i un altre per a les arestes. Aquest, permet definir com ha de ser visualitzat cadascun d'aquests elements.

   (b) Crear i afegir un *renderer* per al text dels nodes i un altre per al text de les arestes. La seva funció és la mateixa que en l'anterior punt.

   (c) Crear i afegir un conjunt d'accions que s'encarreguen d'assignar les propietats pertinents als nodes, arestes i el text que contenen aquests. Les propietats són totes aquelles que no han estat determinades en els *renderers* anteriors.

   (d) Crear i afegir un *layout* que determina la posició de cada node i aresta en la visualització.

   (e) Crear i afegir un conjunt d'accions que permeten fer que la visualització disposi d'animacions i sigui més atractiva.

   (f) Afegir un conjunt de *Listeners* que permeten a la visualització del graf assabentar-se quan l'usuari desitja realitzar alguna acció sobre ella i aplicar-hi els canvis pertinents.

Finalment, la visualització ja ha estat creada, i per tant, es pot afegir la instància de la classe *MyDisplay,* com si fos un component més, a la interfície gràfica de l'aplicació.

5.4.3 Fugues de memòria

És important fer menció d'una mala implementació de la llibreria de *Prefuse [24]* que provocava errors d'espai a memòria. Concretament, el problema rau en què la pila de *Java [18]* s'emplenava molt ràpid sense buidar-se correctament.

Durant el desenvolupament del sistema software, aquest problema no es va fer notar degut a que la implementació, que encara no estava acabada, tan sols permetia visualitzar un graf i calia tornar a executar el programa per visualitzar-ne un altre. Ara bé, quan la codificació ja va ser més avançada i l'aplicació permetia crear noves visualitzacions dels resultats obtinguts en les consultes executades, es va presentar el problema d'espai de memòria a la pila de *Java*.

En un primer moment, es va pensar que l'origen dels errors era el fet que el recol·lector de brossa de *Java* no estava duent a terme la seva funció. Així doncs, es va procedir a executar manualment el recol·lector sense esperar que el propi llenguatge realitzés la tasca d'alliberar memòria. Es va decidir executar el recol·lector en aquells punts que es creia que eren conflictius. El resultat era el mateix, així que va ser necessari esbrinar altres causes i possibles solucions.

Gràcies a l'última versió de l'entorn de desenvolupament integrat (IDE) de *Netbeans [23],* es va poder trobar quin era el causant d'aquests problemàtics errors. El problema es presentava en el moment d'eliminar la visualització. La llibreria de *Prefuse* disposa d'un mètode que s'encarrega d'esborrar tot el què ha estat creat durant la visualització del graf. No obstant, els desenvolupadors d'aquesta llibreria no van tenir en compte que algunes aplicacions que utilitzarien *Prefuse* generarien més d'una visualització durant l'execució del programa.

Qualsevol intent de cercar informació referent a aquest problema o de posar-se en contacte amb desenvolupadors que es trobessin amb el mateix inconvenient va ser un fracàs, així que es va decidir utilitzar algun altre mètode per obtenir una solució. Així que, mitjançant l'anàlisi a temps real de la pila de Java, es va aconseguir trobar les variables que provocaven l'error. El problema doncs, era que moltes de les variables creades en cada visualització no eren eliminades correctament cada vegada que es volia mostrar un nou graf.

Gràcies al fet que aquesta llibreria ofereix públicament el seu codi, ha estat possible reescriure alguns mètodes i afegir-ne de nous per tal de solucionar el problema i garantir la visualització de tants grafs com consultes es desitgin executar.

Capítol 6

Planificació i costos

Després d'haver definit abastament el sistema software desenvolupat és necessari fer esment de tot allò que fa referència a la planificació de les diferents fases del projecte i els costos associats a la seva realització. Tot i què aquest capítol no té res a veure amb l'àmbit de l'enginyeria, podem dir que és de vital importància ja que sense organitzar les tasques d'un procés i determinar-ne la viabilitat és impossible implementar una aplicació en els límits temporals i econòmics establerts pel client.

6.1 Planificació del projecte

La planificació de qualsevol projecte no és una tasca que es realitzi al principi o al final d'un projecte sinó que comença des del primer dia que s'aborda el projecte fins el dia del lliurament final. Descriu l'organització, assignació i duració de les tasques a desenvolupar al llarg del cicle de vida clàssic d'un sistema software.

Cal remarcar la importància de planificar un projecte sempre que existeixi la figura del client involucrada. Aleshores, és quan és més necessari tenir definides les dates per a cada fase amb l'objectiu d'oferir punts de control i presentació dels resultats en aquell punt, a més d'oferir informació de les tasques que resten per finalitzar.

Com a característica especial d'aquest projecte destaca, per una banda, el fet de treballar de cara a un client que fixa els requisits i les dates de lliurament, i per l'altra, que existeixi una limitació de temps marcada pel calendari de la universitat degut al fet que és un projecte final de carrera.

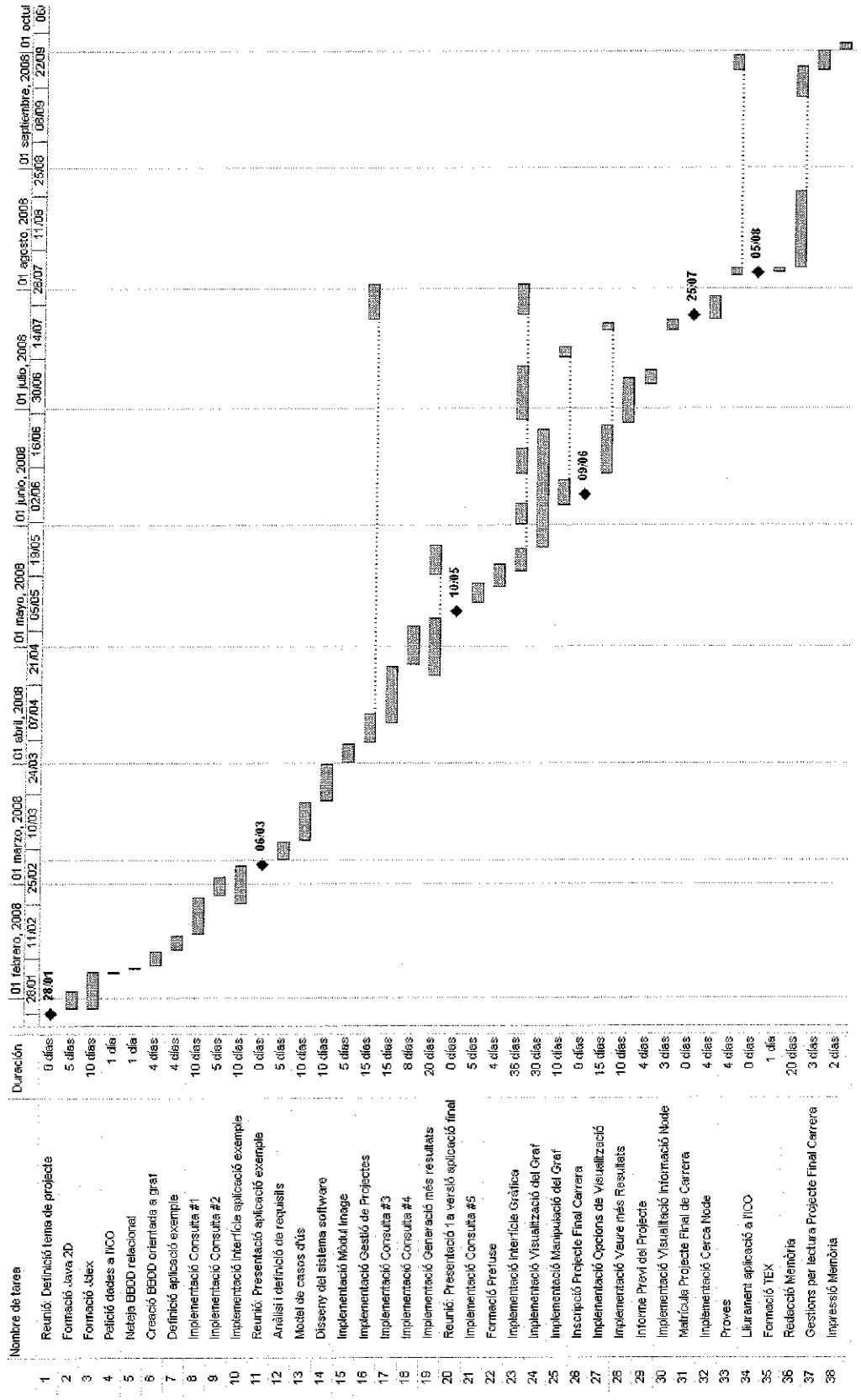
La planificació a l'inici d'un projecte és la base de la previsió de temps per a cada tasca i l'element del qual parteix l'estudi de viabilitat. No obstant, a mesura que passa el temps, cal modificar la planificació realitzada al començament i adaptar-la a les desviacions comeses durant el desenvolupament del projecte amb l'objectiu d'obtenir les hores destinades a cada tasca per tal de calcular-ne els costos finals.

Per tal de simplificar l'extensió del capítol, no es mostrarà la planificació del projecte que es va realitzar al seu inici i el respectiu estudi de viabilitat sinó que, tan sols, s'exposa el resultat de la planificació després d'haver finalitzat el projecte. Cal dir també, que existeixen molts vincles entre tasques que no han estat indicats en el calendari per tal de donar una visió més entenedora de la planificació.

La planificació i els costos han estat desenvolupats i calculats mitjançant la potent eina de gestió de projectes anomenada *Microsoft Project*.

6.1.1 Diagrama de Gantt

El calendari del projecte ha estat realitzat mitjançant un diagrama de gantt atesa la gran facilitat de lectura i seguiment que ofereix aquest tipus de representació. A continuació, es pot observar quina ha estat la durada de cada tasca i els terminis en què s'ha dut a terme.



6.2 Costos del projecte

Tal i com ja s'ha fet èmfasi anteriorment, és molt important dur a terme una planificació a l'inici d'un projecte i un estudi de viabilitat amb l'objectiu de determinar el pressupost i comprovar que es disposen de suficients recursos per tal de fer-ne possible la realització.

Ara bé, quan ha finalitzat un projecte, és necessari anotar els costos finals amb l'objectiu de controlar la quantitat de diners que surten i entren de l'empresa, a més de determinar l'import final del projecte que el client haurà de pagar.

En aquest projecte i en la majoria de projectes, hi intervenen dos tipus de recursos. Per una banda, els recursos humans que s'encarreguen de dur a terme tot el desenvolupament del projecte, i per l'altre, els recursos materials que representen les eines necessàries per a realitzar-ne les tasques.

De totes maneres, cal afegir que en un projecte hi intervenen altres factors que cal indicar en el conjunt de costos totals com poden ser les despeses ocasionades per desplaçaments, imprevistos, etc. que no es tindran en compte en aquest càlcul final.

6.2.1 Costos de personal

Els costos que representen un volum més important del conjunt de despeses d'aquest projecte són els que fan referència a les persones que hi estan involucrades. El fet que aquest projecte és un projecte final de carrera implica que no existeixi un grup encarregat de realitzar les tasques especialitzades sinó que és la figura d'un becari de projectes qui les desenvolupa.

Ara bé, per a mostrar un càlcul dels costos associats al projecte més realista s'ha cregut oportú simular un conjunt de recursos humans inexistents. No obstant, les hores destinades a la realització de cada tasca són totalment verídiques.

6.2.1.1 Recursos humans del projecte

En el desenvolupament d'un sistema software existeixen diverses tasques que formen el cicle de vida clàssic de qualsevol projecte. D'aquesta manera, els encarregats de dur a terme aquestes fases del cicle acostumen a ser un grup d'analistes, dissenyadors, programadors i un cap de projecte.

En aquest projecte en concret, no es té en compte la figura del cap de projecte per tal de simplificar-ne l'assignació de recursos humans i el càlcul de costos de manera que el conjunt de persones involucrades són un analista, un dissenyador, un programador i un formador.

L'analista s'encarrega de la part d'anàlisi i especificació de requisits, el dissenyador realitza el disseny del sistema software, el programador es responsabilitza de la codificació de l'aplicació, i finalment, el formador ensenya les noves tecnologies necessàries per a la realització del projecte.

A continuació, es mostra el barem de preu per hora fixat per a cada treballador.

Nombre del recurso	Tasa estándar
Analista	35,00 €/hora
Dissenyador	30,00 €/hora
Programador	15,00 €/hora
Formador	15,00 €/hora

6.2.1.2 Cost total dels recursos humans del projecte

Un cop es coneix el calendari del projecte, l'assignació de recursos humans a cada tasca i la taxa estàndard per a cada treballador és possible realitzar els càlcul del cost total del projecte.
A continuació es pot veure una taula que s'obté a partir de l'assignació de recursos i la realització del calendari mitjançant l'aplicació *Microsoft Project.*

Id	Nombre de tarea	Nombre del recurso	Trabajo	Costo total
36	Redacció Memòria	Analista[50%];Dissenyador[50%]	80 horas+80 horas	5.200,00 €
23	Implementació Interfície Gràfica	Programador	288 horas	4.320,00 €
24	Implementació Visualització del Graf	Programador	240 horas	3.600,00 €
13	Model de casos d'ús	Analista	80 horas	2.800,00 €
14	Disseny del sistema software	Dissenyador	80 horas	2.400,00 €
19	Implementació Generació més resultats	Programador	160 horas	2.400,00 €
16	Implementació Gestió de Projectes	Programador	120 horas	1.800,00 €
17	Implementació Consulta #3	Programador	120 horas	1.800,00 €
27	Implementació Opcions de Visualització	Programador	120 horas	1.800,00 €
12	Anàlisi i definició de requisits	Analista	40 horas	1.400,00 €
3	Formació Jdex	Formador	80 horas	1.200,00 €
8	Implementació Consulta #1	Programador	80 horas	1.200,00 €
10	Implementació interfície aplicació exemple	Programador	80 horas	1.200,00 €
25	Implementació Manipulació del Graf	Programador	80 horas	1.200,00 €
28	Implementació Veure més Resultats	Programador	80 horas	1.200,00 €
7	Definició aplicació exemple	Analista;Dissenyador	16 horas+16 horas	1.040,00 €
29	Informe Previ del Projecte	Analista[50%];Dissenyador[50%]	10 horas+20 horas	950,00 €
18	Implementació Consulta #4	Programador	64 horas	960,00 €
2	Formació Java 2D	Formador	40 horas	600,00 €
9	Implementació Consulta #2	Programador	40 horas	600,00 €
15	Implementació Mòdul Image	Programador	40 horas	600,00 €
21	Implementació Consulta #5	Programador	40 horas	600,00 €
6	Creació BBDD orientada a graf	Programador	32 horas	480,00 €
22	Formació Prefuse	Formador	32 horas	480,00 €
32	Implementació Cerca Node	Programador	32 horas	480,00 €
33	Proves	Programador	32 horas	480,00 €
30	Implementació Visualitació Informació Node	Programador	24 horas	360,00 €
5	Neteja BBDD relacional	Programador	8 horas	120,00 €
35	Formació TEX	Formador	8 horas	120,00 €
1	Reunió: Definició tema de projecte	Analista;Dissenyador;Programador		0,00 €
4	Petició dades a l'ICO	Analista;Dissenyador;Programador		0,00 €
11	Reunió: Presentació aplicació exemple	Analista;Dissenyador;Programador		0,00 €
20	Reunió: Presentació 1a versió aplicació final	Analista;Dissenyador;Programador		0,00 €
26	Inscripció Projecte Final Carrera			0,00 €
31	Matrícula Projecte Final de Carrera			0,00 €
34	Lliurament aplicació a l'ICO	Programador		0,00 €
37	Gestions per lectura Projecte Final Carrera			0,00 €
38	Impressió Memòria			0,00 €
			2.262 horas	41.390,00 €

Es pot veure doncs, que el total de temps destinat a la *realització* del projecte ha estat de 2.262 hores i el cost total del conjunt de les tasques és de 41.390 euros.

6.2.2 Costos de material

Finalment, cal tenir en compte els costos derivats de la utilització de maquinari i software per al desenvolupament del projecte.

A continuació, es mostra el conjunt de software que requereix el pagament d'una llicència, o bé, aquells elements del maquinari que han hagut de ser comprats durant la realització del projecte. Es calcula que són necessaris tres equips informàtics amb els respectius perifèrics per tal de realitzar el sistema software. A més, pel que fa les aplicacions que requereixen el pagament d'una llicència, s'ha utilitzat el sistema operatiu *Microsoft* a *XP* i el paquet *Microsoft Office 2003* que conté l'aplicació Microsoft Project 2003 la qual permet realitzar planificacions i càlculs de costos.

Altres programes utilitzats com l'editor de text en format latex anomenat *Lyx* [22], l'eina de desenvolupament *Netbeans 6.0* [23], etc. són de lliure distribució i, per tant, no influeixen en els costos de material.

Equips informàtics	3.000 €
Microsoft Windows XP	125 €
Microsoft Office 2003	50 €
TOTAL	3.175 €

6.2.3 Cost total

Finalment doncs, podem dir que el total dels costos del projecte és la suma dels costos dels recursos humans i els costos de material.

COST TOTAL	41.390 € + 3.175 € = 44.565 €

Capítol 7

Conclusions

En aquest capítol es comenten tots aquells aspectes que han contribuït en el desenvolupament del projecte i tot el què ha aportat a nivell personal i professional.

7.1 Implicació personal

En aquest projecte cal destacar la figura d'un membre de la família, que alhora és codirectora d'aquest projecte, la qual exerceix d'investigadora en l'àmbit de l'oncologia. Aquesta doncs, és la motivació principal a l'hora d'elegir el tema del projecte. Durant anys, el seu treball en aquest àmbit ha influït en despertar l'interès de la projectista en el desenvolupament d'un sistema software per a la investigació en medicina. A més, ha estat molt més fàcil i ràpida la tasca d'entendre el context del projecte.

Un altre dels motius principals pel qual es realitza aquest projecte és el fet d'aportar una eina per a l'avanç en el coneixement sobre la utilització dels serveis en els centres sanitaris i dels tractaments oncològics a Catalunya de manera objectiva. Cal recordar que el càncer és, actualment, la segona causa de mort després de les malalties del cor. De totes maneres, es preveu que al llarg d'aquest segle, n'esdevingui la primera causa de mort. També cal dir que s'estima que augmenti la supervivència dels pacients amb malalties del càncer al llarg d'aquest segle. Per tant, el nombre de persones diagnosticades de càncer i ingressades en algun centre sanitari augmentarà gradualment. És evident doncs, que cal treballar dur en aquest terreny de la medicina per tal de contribuir a quantificar d'una forma clara, estructurada i objectiva com s'utilitzen els recursos sanitaris en els processos assistencials en aquests pacients. Aquesta és la informació necessària per al desenvolupament de la planificació sanitària que garanteixi l'èxit de la utilització dels serveis sanitaris.

Cal mencionar l'interès personal en el fet d'aconseguir un lloc com a becària de projectes al grup DAMA-UPC [14]. El departament d'Arquitectura de Computadors (AC) de la Universitat Politècnica de Catalunya (UPC) sempre ha estat conegut per la seva professionalitat i gran coneixement de la matèria. Treballar dins un grup d'aquest departament doncs, permetia disposar d'una gran oportunitat per aprendre'n del seu coneixement i experiència.

D'altra banda, també es va considerar interessant treballar per a un client amb l'objectiu d'aprendre a saber transmetre i entendre els missatges que en garanteixen una bona comunicació. Tot i que a primera vista sembla una tasca senzilla, a l'hora de la veritat es pot convertir en l'obstacle principal que impedeix avançar en l'especificació d'un sistema software. Cal saber extreure, del professional mèdic, la informació necessària per tal de construir-ne els requisits del sistema i especificar-los. Un investigador, sovint no té coneixements informàtics com tampoc en té de medicina un desenvolupar de sistemes software. Ha estat un dur treball doncs, entendre quin era l'objectiu del client i saber-ne construir el camí entre ambdós parts.

7.2 Coneixements previs

Tot i que el departament d'Arquitectura de Computadors es centra principalment en temes relacionats amb el hardware, cal dir que el grup DAMA-UPC disposa d'un conjunt de gent que es dedica al desenvolupament de software. Concretament, DAMA-UPC centra la seva investigació en l'eficiència i èxit en el tractament de grans volums de dades.

Tal i com s'ha dit abans, existia un interès pel desenvolupament d'una eina destinada a la investigació del càncer. La manipulació de dades en Bases de Dades de grans dimensions com són els registres i Bases de Dades d'un centre sanitari fa que no sigui senzill extreure'n informació de manera ràpida i fiable.

D'aquesta manera, a part de les assignatures obligatòries cursades que han permès desenvolupar aquest projecte, han estat útils els coneixements aportats per algunes assignatures optatives que amplien el coneixement necessari per a tractar grans volums de dades.

Disseny i administració de Bases de Dades (DABD): En aquesta assignatura s'adquireixen coneixements que introdueixen a l'alumne al món de les Bases de Dades i, per tant, era important cursar-la per a entendre la dificultat del tractament amb moltes dades i tots els problemes que podem trobar-nos davant el disseny i administració d'una Base de Dades. Cal recordar que, en aquest projecte, es realitza una transformació d'un model de dades relacional a un model de dades orient a graf.

Mineria de dades (MD): En aquesta assignatura es recullen tots aquells aspectes que fan referència a l'assignació de patrons per cada conjunt de dades seleccionat. En aquest projecte l'objectiu és obtenir uns resultats a partir de la manipulació d'unes dades determinades per tal de generar regles i algorismes. Així doncs, el coneixement aportat per aquesta assignatura ha permès entendre millor quina informació s'ha de mostrar que faciliti la identificació i determinació d'aquestes regles.

Recuperació de la Informació (RI): En aquesta assignatura s'introdueixen conceptes de com recuperar informació emmagatzemada en grans Bases de Dades. En aquest projecte doncs, també es treballa amb grans volums de dades i s'intenta construir una informació a partir de la recuperació de dades. També s'hi introdueixen conceptes que fan referència a la mineria de dades.

Cal tenir en compte altres assignatures optatives que també han ajudat en la redacció de la memòria final com les que s'exposen a continuació:

Planificació i Gestió de Projectes i Sistemes Informàtics (PGPSI): S'hi adquireixen coneixements que ajuden a saber construir el calendari d'un projecte, a més de mostrar tècniques i tecnologia per a poder-lo gestionar.

Viabilitat de Projectes Empresarials (VPE): L'objectiu d'aquesta assignatura és la redacció d'un pla de viabilitat d'una empresa fictícia inventada per estudiants. En aquest pla, cal indicar-hi aspectes com els costos, les planificacions, el màrqueting, etc. La realització d'aquesta doncs, ha permès redactar el capítol de planificació.

7.3 Coneixements adquirits

Després d'haver finalitzat el projecte són molts els coneixements que s'han adquirit. No obstant, cal dir que quasi tots són de caràcter pràctic tal i com pretén la realització d'un projecte final de carrera.

Així doncs, a continuació es llista un resum d'aquests coneixements:

- Introducció a la investigació en oncologia.
- Experiència en la comunicació entre client i desenvolupador.
- Implementació de software mitjançant el llenguatge *Java*.
- Utilització de l'eina d'anàlisi d'ús de memòria de *Netbeans 6.0*.
- Utilització de la llibreria *Jdex* de DAMA-UPC.
- Utilització de la llibreria *Jakarta Commons Configuration* per a la implementació de configuracions.
- Disseny d'interfícies gràfiques mitjançant *Swing*.
- Utilització de la llibreria *JExcelApi* per la construcció i lectura de fitxers de *Microsoft Excel*.
- Utilització de la llibreria de *Prefuse* per la visualització de graf.
- Redacció de documents mitjançant LaTeX.

Capítol 8

Resultats i Futur del Software

Després d'haver finalitzat el desenvolupament del software i haver-ne lliurat l'aplicació final al client, és interessant veure quins resultats s'obtenen a partir de la utilització d'aquesta nova eina. A més, existeix un projecte que permet continuar amb l'ampliació de l'eina desenvolupada.

8.1 Període de proves

Tot i que l'aplicació ha estat implementada seguint les funcionalitats requerides pel client i les definicions de tot allò pertanyent a l'àmbit de la medicina que resultava inabastable pel grup DAMA-UPC [17], és evident que existeixen errors de codificació.

Per tant, ha estat necessària la comprovació de la veridicitat dels resultats obtinguts en les execucions de consultes que ofereix l'eina realitzada. Aquesta tasca, l'ha dut a terme el grup d'investigació que representa el rol de client en aquest projecte. S'han hagut de comparar els resultats obtinguts en l'execució de les diferents consultes amb el coneixement dels investigadors que, durant anys, han adquirit suficient experiència per a determinar si aquesta informació resultant és apta per a l'anàlisi i extracció de regles i algorismes per a cada tumor. Cal dir que, en la majoria dels casos, aquesta comprovació ha estat tan sols una tasca per assegurar-ne la veridicitat sense que hagi estat necessari realitzar cap canvi en la implementació. Per a la validació de l'eina, també hi han treballat els codificadors del grup amb l'objectiu de verificar que els codis que s'exposen en els resultats són correctes i no s'ha produït cap desviació d'allò que van demanar amb el què se'ls hi ha ofert finalment.

8.2 Anàlisi de resultats

Després d'haver dut a terme el període de proves, el grup d'investigació que actua com a client en aquest projecte ha començat l'anàlisi de tots aquells possibles resultats que s'obtenen a partir d'una localització tumoral. Després de validar-ne l'eina, el següent pas que han de dur a terme els investigadors és aprendre a utilitzar l'aplicació i saber-ne treure el màxim profit amb l'objectiu d'extreure'n els resultats necessaris per tal de dur a terme un estudi complert.

Cal recordar que l'objectiu del desenvolupament d'aquest sistema software no era un altre que el d'obtenir uns resultats que permetessin avaluar els diferents procediments terapèutics i diagnòstics implicats en els ingressos hospitalàries de les persones diagnosticades d'un tumor en concret. Aquesta avaluació ha de permetre als investigadors la generació d'unes regles i algorismes que seran introduïdes a l'aplicació ASEDAT [24] amb l'objectiu d'obtenir automàticament unes dades específiques de la Base de Dades de les AH que

fan referència a procediments diagnòstics i terapèutics. L'extracció automàtica d'aquestes dades permetrà construir el RHT que ha de contenir la informació referent a les persones diagnosticades d'un tumor juntament amb els procediments implicats en els ingressos posteriors al diagnòstic d'aquest tumor. L'objectiu de les dades contingudes en aquest nou RHT és facilitar l'elaboració d'una planificació sanitària que garanteixi l'èxit en la utilització dels serveis sanitaris destinats als processos assistencials dins de cada centre sanitari de Catalunya.

Actualment, els investigadors del grup, estan executant les consultes que creuen oportunes per a cada localització tumoral amb l'objectiu d'extreure'n les regles i algorismes necessàries. El següent pas doncs, és construir-ne aquestes regles per a introduir-les a l'ASEDAT i obtenir automàticament el RHT amb les dades adients per a realitzar la planificació sanitària.

8.3 Futur del software

El fet que un projecte acadèmic tingui uns límits temporals i acadèmics fixats impedeix estendre's més en la seva implementació. Ara bé, en aquest cas, existeix un projecte de continuació i ampliació d'aquest sistema software. És per aquest motiu que, a continuació, s'exposa com prosseguirà i en quins altres àmbits de la medicina pot aplicar-se.

8.3.1 Continuïtat del projecte

Tot i que l'aplicació implementada està sent utilitzada pel grup d'investigació que forma part de l'ICO, és possible i necessari realitzar ampliacions i oferir-ne suport.

L'aplicació final disposa de cinc consultes diferents que poden ser executades per a cada localització tumoral. Els resultats que ofereixen aquestes consultes permeten realitzar una anàlisi del comportament de cada tumor. Ara bé, a mesura que aquest grup d'investigació vagi adquirint experiència en l'ús de l'aplicació, sorgiran noves idees a tenir en compte per tal de millorar-ne les funcionalitats.

A curt termini, podem dir que serà necessari dur a terme una modificació d'aquella part del codi que no s'hagi implementat correctament i desemboqui en resultats erronis, o bé, oferir la possibilitat d'afegir noves consultes que aportin informació útil i interessant per a l'anàlisi de la informació associada a cada tumor. De fet, s'ha arribat a un acord entre el client i el grup DAMA-UPC per a continuar en el desenvolupament de l'aplicació per tal de dur a terme les ampliacions que siguin necessàries o es cregui convenient realitzar.

8.3.2 Possibles ampliacions

Tot i què encara no es coneixen els aspectes que cal millorar de l'aplicació desenvolupada, a continuació s'ofereix una llista amb aquelles possibles ampliacions que podrien ser realitzades en un futur pròxim:

- Ampliació de la informació que s'obté en els resultats obtinguts en l'execució de consultes.
- Ampliació de la quantitat de diferents formats de representació pels resultats obtinguts en l'execució de consultes.
- Visualització més atractiva dels fitxers de resultats addicionals en diferents formats obtinguts en l'execució de consultes.
- Ampliació de la quantitat i qualitat de les possibles consultes a executar.

- Visualització més atractiva i/o interactiva del graf resultant de l'execució de consultes.
- Possibilitat d'escollir el destí on s'emmagatzema cada projecte i la informació associada.
- Eines de filtratge de la visualització del graf més potents.
- Suport en diferents idiomes.
- Inclusió d'altres variables a la Base de Dades relacional per a executar consultes més diverses.
- Desenvolupament d'un importador que realitzi la tasca de creació de la Base de Dades relacional a partir d'un registre de diagnòstics i una Base de Dades de les AH.

Tal i com s'ha explicat abans, el fet que l'aplicació es desenvolupi com a projecte final de carrera no ha permès que s'aprofundís en el desenvolupament dels anteriors aspectes llistats per les restriccions de temps i càrrega de treball. Tanmateix, l'aplicació final que es presenta és un sistema software complert i preparat per a ser utilitzat de manera que compleix amb les expectatives fixades pel client en l'inici del projecte.

8.3.3 Aplicació en altres especialitats

En el capítol on s'introduïa el context del projecte s'han definit què era el RHT [25], el Registre de Càncer de Catalunya i, finalment, la Base de Dades de les AH. L'aplicació desenvolupada en aquest projecte ha procurat partir d'una Base de Dades relacional que manté el format especificat pel registre del CMBD [16]. Aquesta característica permet l'aplicació del software en altres especialitats. Tan sols és necessari que cada especialitat mèdica disposi d'un registre que contingui els diagnòstics corresponents a una patologia en concret relacionada amb l'especialitat de la mateixa manera que, en aquest cas, es disposa del RHT que conté els diagnòstics de tumors. Pel què fa les AH, es dóna per fet que cada centre sanitari disposa d'aquest registre que conté el total d'ingressos hospitalaris registrats en aquell mateix centre. A més, cada Base de Dades de les AH de qualsevol centre ha de seguir l'especificació marcada pel registre del CMBD.

D'aquesta manera, partint d'un registre de diagnòstics d'una patologia i de la Base de Dades de les AH d'un o més centres sanitaris, es pot crear la Base de Dades relacional que actua com a element d'entrada en l'aplicació desenvolupada. Per a què el sistema software funcioni correctament, cal recordar que aquesta Base de Dades ha de complir amb l'especificació descrita en els capítols anteriors. Actualment, l'eina no permet crear la Base de Dades relacional de manera automàtica sinó que caldria implementar un importador amb l'objectiu de carregar les dades del registre de diagnòstics i les del les AH per tal de construir-ne la Base de Dades relacional necessària per executar l'aplicació.

Després d'haver construït la Base de Dades relacional, ja estaria tot a punt per a dur a terme l'execució de les cinc consultes que ofereix el sistema software desenvolupat. Els resultats obtinguts serien fàcilment interpretables per un expert de l'especialitat mèdica en qüestió, de la mateixa manera que es realitza actualment per a l'especialitat d'oncologia.

A continuació, s'ofereix un exemple per a entendre-ho més clarament. L'especialitat de cardiologia també disposa d'un gran volum de dades emmagatzemades en registres específics. Així doncs, la investigació d'aquesta àrea de la medicina també requereix d'eines informàtiques que en facilitin l'extracció d'informació amb l'objectiu d'estudiar-ne el comportament de cadascuna de les malalties cardiovasculars que es coneixen. Així doncs, tan sols seria necessari prendre un registre que emmagatzemés els diagnòstics de malalties cardiovasculars i una Base de Dades de les AH d'un o més centres sanitaris. Consegüentment, cada projecte partiria d'un diagnòstic de malaltia del cor enlloc de fer-ho d'una localització tumoral. Els resultats de les

diverses execucions de cada consulta oferirien un conjunt d'informació útil tal i com ho fa actualment el sistema software desenvolupat amb l'estudi de les malalties oncològiques.

És evident doncs, que aquest software és fàcilment aplicable a qualsevol especialitat mèdica, la qual cosa permet estendre'n l'ús amb la finalitat de col·laborar amb l'avanç en la recerca mèdica.

Apèndix A

Fitxers de resultats

En aquest capítol es descriu quina és la representació per a cada fitxer que es pot generar en l'execució d'una consulta. Existeix un total de cinc fitxers diferents que contenen la mateixa informació representada en diferents formats per oferir un punt de vista més ampli. Cal recordar que aquests fitxers contenen la informació resultant que s'obté en cadascuna de les cinc consultes que ofereix l'aplicació.

A.1 Introducció

Cal recordar que és necessari que, cada cop que s'executi una de les cinc consultes, l'usuari indiqui quina informació addicional desitja que l'aplicació generi. Així doncs, hi ha cinc possibles fitxers a construir on s'hi emmagatzema la informació pertinent.

De totes maneres, és evident que cada fitxer que s'explica a continuació conté una informació o una altra depenent de quina de les cinc consultes s'executi.

A.2 Fitxer de resultats en format textual

La informació continguda en aquest fitxer és la mateixa que es mostra en la visualització del graf resultant de cada execució de consulta. L'avantatge principal és la possibilitat d'imprimir el contingut ja que, d'aquesta manera, permet estudiar una part dels resultats detingudament sense necessitat de navegar per la visualització del graf.

Tal i com es pot veure en els següents gràfics, aquests fitxers mostren la informació de manera que cada nivell de nodes en la visualització del graf es distingeix gràcies a tabulacions del text. Per tant, cada nova tabulació és un nou nivell de profunditat dins el graf.

Una altra manera de representar-ho és utilitzant claudàtors de manera que, cada nivell de nodes que representa un tipus d'informació diferent, apareix escrit entre claus. Alhora, els claudàtors simbolitzen un node de la visualització del graf, per tant, tot el que hi ha dins el fragment de text que va des de l'inici d'aquesta clau [NOM] fins al final [/NOM] és la informació que conté el propi node juntament amb tota la informació dels seus fills representants mitjançant claus. El text de color taronja assenyala aquells fragments que varien segons les dades d'entrada.

A continuació es mostren els gràfics que permeten observar el format per a cadascun dels possibles fitxers generats per cada consulta. Atès que, en altres capítols, s'ha dut a terme una definició exhaustiva de les consultes, no es comentaran individualment.

```
[TUMOR INCIDENT] CODI ICD9 DEL TUMOR: DESCRICIÓ DEL TUMOR
    * total procediments aplicats = X
    [PROC] CODI PROCEDIMENT - DESCRIPCIÓ PROCEDIMENT
        * total procediments aplicats = R
        * % absolut = S
        * id. node = T
    [/PROC]
    [PROC] CODI PROCEDIMENT - DESCRIPCIÓ PROCEDIMENT
        * total procediments aplicats = X
        * % absolut = Y
        id. node = Z
    [/PROC]
    ...
[TUMOR INCIDENT]
```

Figura A.1: Estructura del fitxer de resultats en format textual de la consulta 1

```
[TUMOR INCIDENT] CODI ICD9 DEL TUMOR: DESCRICIÓ DEL TUMOR
    * total procediments aplicats = X
    [PROC] CODI PROCEDIMENT - DESCRIPCIÓ PROCEDIMENT
        * total procediments aplicats = R
        * % absolut = S
        * id. node = T
    [/PROC]
    [PROC] CODI PROCEDIMENT - DESCRIPCIÓ PROCEDIMENT
        * total procediments aplicats = X
        * % absolut = Y
        id. node = Z
    [/PROC]
    ...
[TUMOR INCIDENT]
```

Figura A.2: Estructura del fitxer de resultats en format textual de la consulta 2

```
[TUMOR INCIDENT] CODI TUMOR INCIDENT: DESCRIPCIÓ TUMOR INCIDENT
    * total incidències = A
    [SENSE AH INGRES]
        * total ingressos = B
        * % absolut = C
        * id. node = D
    [/SENSE AH  INGRES]
    [1 AH INGRES]
        * total ingressos = E
        * % absolut = F
        * id. node = G
        [PROC PRINC] CODI PROCEDIMENT PRINCIPAL SENSE DETALL: DESCRIPCIÓ
            * repeticions =  H
            * % absolut = I
            * % relatiu = J
            * id. node = K
            [PROC PRINC - molt detall] CODI PROCEDIMENT PRINCIPAL AMB DETALL: DESCRIPCIÓ
                * repeticions = L
                * % absolut = M
                * % relatiu = N
                * id. node = O
            [/PROC PRINC - molt detall]
            [PROC PRINC - molt detall] CODI PROCEDIMENT PRINCIPAL AMBDETALL: DESCRIPCIÓ
                ...
            [/PROC PRINC - molt detall]
            [NO 2 AH INGRES]
              * re-ingressos = P
              * % absolut = Q
              * % relatiu = R
              * id. node = S
            [/NO 2 AH INGRES]
            [2 AH INGRES]
                ...
            [/2   AH INGRES]
                ...
```

```
            [N AH INGRES]
                ...
            [N-1 AH INGRES]
        [/PROC PRINC]
    [/1 AH INGRES]
[/TUMOR INCIDENT]
```

Figura A.3: Estructura del fitxer de resultats en format textual de la consulta 3

```
[TUMOR INCIDENT] CODI TUMOR INCIDENT: DESCRIPCIÓ TUMOR INCIDENT
    * total incidències = A
    [SENSE AH INGRES]
        * total ingressos = B
        * % absolut = C
        * id. node = D
    [/SENSE AH INGRES]
    [1 AH INGRES]
        * total ingressos = E
        * % absolut = F
        * id. node = G
        [DIAG PRINC] CODI DIAGNÒSTIC PRINCIPAL: DESCRIPCIÓ
            * repeticions = X
            * % absolut = Y
            * % relatiu = Z
            * id. node = T
            [PROC PRINC] CODI PROCEDIMENT PRINCIPAL SENSE DETALL: DESCRIPCIÓ
                * repeticions = H
                * % absolut = I
                * % relatiu = J
                * id. node = K
                [PROC PRINC - molt detall] CODI PROCEDIMENT PRINCIPAL AMB DETALL: DESCRIPCIÓ
                    * repeticions = L
                    * % absolut = M
                    * % relatiu = N
                    * id. node = O
                [/PROC PRINC - molt detall]
                [PROC PRINC - molt detall] CODI PROCEDIMENT PRINCIPAL AMBDETALL: DESCRIPCIÓ
                    ...
                [/PROC PRINC - molt detall]
            [/PROC PRINC]
        [/DIAG PRINC]
    [/1 AH INGRES]
[/TUMOR INCIDENT]
```

Figura A.4: Estructura del fitxer de resultats en format textual de la consulta 4

```
[TUMOR INCIDENT] CODI TUMOR INCIDENT: DESCRIPCIÓ TUMOR INCIDENT
    * total incidències = A
    [SENSE AH INGRES]
        * total ingressos = B
        * % absolut = C
        * id. node = D
    [/SENSE AH INGRES]
    [1 AH INGRES]
        * total ingressos = E
        * % absolut = F
        * id. node = G
        [DIAG PRINC] CODI DIAGNÒSTIC PRINCIPAL: DESCRIPCIÓ
            * repeticions = X
            * % absolut = Y
            * % relatiu = Z
            * id. node = T
            [PROC PRINC] CODI PROCEDIMENT PRINCIPAL SENSE DETALL: DESCRIPCIÓ
                * repeticions = H
                * % absolut = I
                * % relatiu = J
                * id. node = K
```

```
[PROC PRINC - molt detall] CODI PROCEDIMENT PRINCIPAL AMB DETALL: DESCRIPCIÓ
    * repeticions = L
    * % absolut = M
    * % relatiu = N
    * id. node = O
[/PROC PRINC - molt detall]
[PROC PRINC - molt detall] CODI PROCEDIMENT PRINCIPAL AMBDETALL: DESCRIPCIÓ
    ...
[/PROC PRINC - molt detall]
[NO 2 AH INGRES]
  * re-ingressos = P
  * % absolut = Q
  * % relatiu = R
  * id. node = S
[/NO 2 AH INGRES]
[2 AH INGRES]
    ...
[/2  AH INGRES]
...
[N AH INGRES]
    ...
[N-1 AH INGRES]
[/PROC PRINC]
[/DIAG PRINC]
[/1 AH INGRES]
[/TUMOR INCIDENT]
```

Figura A.5: Estructura del fitxer de resultats en format textual de la consulta 5

A.3 Fitxer de resultats de Microsoft Excel

Els fitxers de Microsoft Excel, de la mateixa manera que ho feien els fitxers de resultats en format textual, representen la mateixa informació que la visualització del graf. La característica més important és l'avantatge que ofereix aquest format ja que permet disposar de les dades en fulls de càlcul que organitzen la informació en taules. A més, Microsoft Excel permet exportar les dades en altres formats que en poden facilitar l'estudi.

A continuació es mostra el format d'aquests fitxers depenent de la consulta que els generi. No es durà a terme una descripció extensa dels resultats de cada consulta ja que s'han definit prèviament els objectius de cadascuna d'elles.

La primera consulta tan sols representa els diferents procediments i el percentatge que representen sobre el total.

TOTES AH		
Procediments	% absolut	repeticions
CODI		
. . .		
CODI		

Figura A.6: Estructura del fitxer de resultats de Microsoft Excel de la consulta 1

A la segona consulta, es representa pràcticament la mateixa informació que a la primera. L'única diferència és que ara són diagnòstics enlloc de procediments.

TOTES AH		
Diagnòstics	% absolut	repeticions
CODI		
...		
CODI		

Figura A.7: Estructura del fitxer de resultats de Microsoft Excel de la consulta 2

La tercera, quarta i cinquena consulta generen un fitxer on les files de la taula representen tots els possibles camins del graf des de l'arrel fins a cada node. L'objectiu és representar la informació continguda dins de cada node juntament amb la informació de tots els nodes que el precedeixen.

1 AH		2 AH			N AH			
Proc sense detall	repeticions	Proc amb detall	repeticions	...	Proc amb detall	% absolut	% relatiu	repeticions
CODI		CODI		...	CODI			
...								
CODI		CODI		...	CODI			

Figura A.8: Estructura del fitxer de resultats de Microsoft Excel de la consulta 4

1 AH		2 AH			N AH			
Proc sense detall	repeticions	Proc amb detall	repeticions	...	Proc amb detall	% absolut	% relatiu	repeticions
CODI		CODI		...	CODI			
...								
CODI		CODI		...	CODI			

Figura A.9: Estructura del fitxer de resultats de Microsoft Excel de la consulta 5

1 AH		2 AH			N AH										
DP	rep	DP	rep	...	DP	%abs	%rel	rep	P detall	%abs	%rel	rep	P no detall	%abs	%rel	rep
CODI		CODI		...	CODI				CODI							
...																
CODI		CODI		...	CODI				CODI							

Figura A.10: Estructura del fitxer de resultats de Microsoft Excel de la consulta 5

A.4 Fitxer de resultats en format XML

Existeix la possibilitat de generar un fitxer que contingui tota la informació representada en la visualització del graf mitjançant el format XML. És a dir, l'aplicació s'encarrega de construir aquest fitxer introduint-hi

la informació de tal manera que el resultat és un fitxer XML organitzat amb etiquetes, les quals representen els diferents elements del graf.

Per a entendre-ho més bé, cal recordar quina era l'estructura del fitxer de resultats en format textual. Cada node quedava representat dins d'una etiqueta envoltada de dos claudàtors. Així doncs, XML és un format que segueix una estructura similar a la presentada en els resultats en format textual de manera que la informació es desa seguint una estructura de graf. Alhora, permet emmagatzemar la informació de tal manera que, més endavant, es pot visualitzar mitjançant una aplicació que permeti la lectura d'aquest fitxer i la representació de la informació en forma de graf. Existeixen aplicacions que ho permeten.

A.5 Llista de pacients en format textual

Aquest fitxer conté el conjunt de pacients agrupats en cadascun dels nodes representats en la visualització del graf resultant. L'avantatge de disposar d'aquesta informació és la possibilitat d'incorporar fàcilment aquestes dades a la Base de Dades Relacional i, per tant, obtenir ràpidament tota la informació que fa referència a cada pacient agrupat en un node determinat.

Així doncs, durant l'anàlisi i estudi de la informació obtinguda en cada consulta, és possible obtenir tots els pacients representats en un node en concret amb l'objectiu de facilitar-ne la interpretació dels resultats.

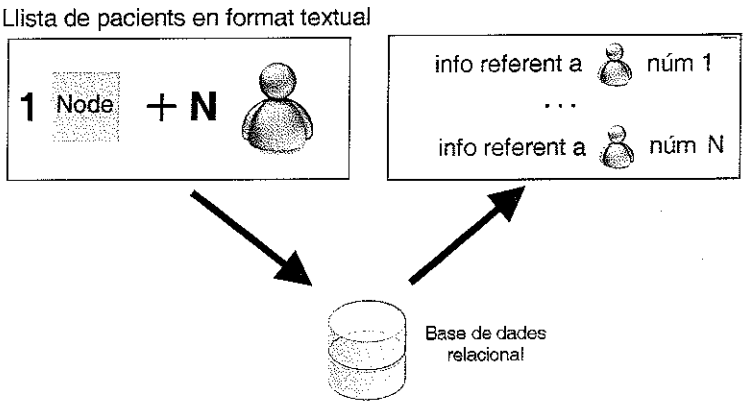


Figura A.11: Procés d'obtenció d'informació dels pacients relacionats amb un node

Finalment, a la Figura A.12 es presenta el format que segueix aquest fitxer.

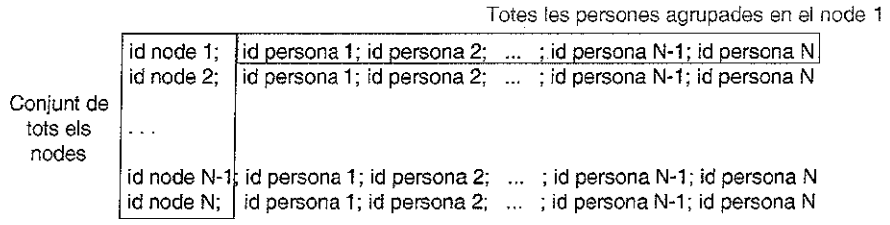


Figura A.12: Estructura de la llista de pacients en format textual

A.6 Llista d'ingressos en format textual

L'estructura d'aquest fitxer és pràcticament igual que l'anterior. L'única diferència és el fet que, en aquest cas, es disposa de la llista dels ingressos representats en cada node enlloc de la llista de persones.

Per tant, l'avantatge també és el mateix. En aquest cas, enlloc d'obtenir la informació de tots els pacients relacionats amb un node mitjançant les dades de la Base de Dades relacional s'obté la informació que fa referència als ingressos.

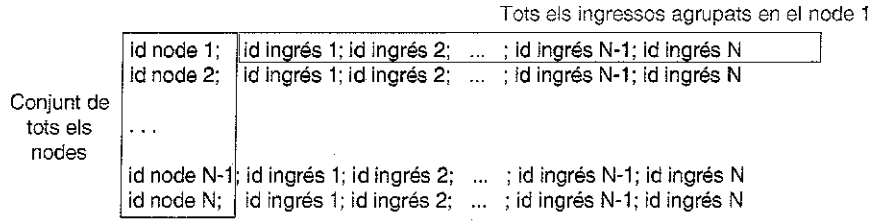


Figura A.13: Estructura de la llista d'ingressos en format textual

Apèndix B

Manual d'usuari

B.1 Introducció

És una eina que permet analitzar registres hospitalaris de càncer amb l'objectiu d'obtenir-ne diversos comportaments i tendències.

B.2 Requisits del sistema

Tot i què l'aplicació pot ser executada en entorns inferiors, es recomana que el sistema compleixi amb els següents requisits de hardware i software:

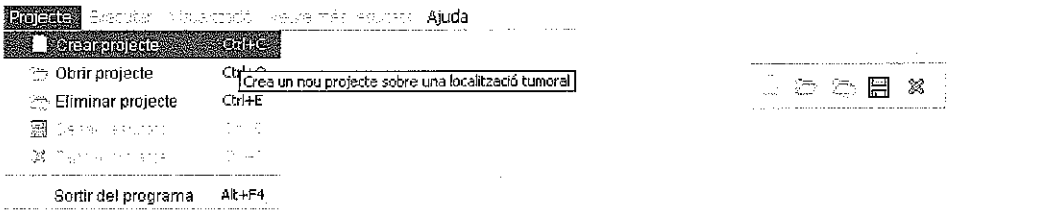
Sistema operatiu	Microsoft Windows XP
Editor de full de càlcul	Microsoft Excel
Freqüència del processador	1 GHz o més
Memòria RAM	512 MB o més
Espai mínim a disc	300 MB

B.3 Gestió de projectes

Un projecte té assignada una localització tumoral a partir de la qual s'executaran les diferents consultes amb l'objectiu d'obtenir la informació necessària per realitzar una anàlisi del comportament relacionat amb aquell tumor. La gestió de projectes ofereix el següent conjunt de funcionalitats:

Crear projecte	Crear un nou projecte i assignar la localització tumoral escollida per l'usuari
Obrir projecte	Obrir un projecte creat amb anterioritat
Eliminar projecte	Eliminar un projecte creat amb anterioritat
Desar resultats	Desar els fitxers generats a petició de l'usuari que contenen els resultats obtinguts en la consulta executada
Tancar projecte	Tancar tot el què fa referència al projecte i als fitxers i visualització generats en les consultes

Totes elles es poden trobar en el menú que porta el nom 'Projecte', o bé, a la barra d'eines de gestió de projectes.



(a) Menú 'Projecte' (b) Barra d'eines de gestió de projectes

Figura B.1: Opcions de gestió de projectes

B.3.1 Crear projecte

Per tal de crear un nou projecte i assignar-li una localització tumoral, seleccioni l'opció del menú 'Projecte', i dins d'aquest premi, a 'Crear projecte', o bé, seleccioni la primera opció de la barra d'eines de gestió de projectes.

A continuació, li apareixerà una finestra on l'aplicació li demana que introdueixi el nom amb el qual vol emmagatzemar a disc el nou projecte. Recordi que cada cop que creï un nou projecte, aquest serà desat al directori on tingui instal·lada l'aplicació.

Un cop hagi escrit el nom del projecte pot prémer el botó 'Accepta'. Acte seguit, apareixerà una nova finestra on l'aplicació li demana que indiqui el codi de la localització tumoral que desitja assignar al projecte. Cal seleccionar-lo i prémer el botó 'Accepta' per a finalitzar el procés de creació del projecte. Per a obtenir la descripció de la localització tumoral que desitja escollir, tan sols ha de posar el cursor del ratolí damunt del codi i es mostrarà un panell informatiu.

(a) Introduir el nom del projecte (b) Seleccionar la localització tumoral (c) Descripció del codi

Figura B.2: Procés per a crear un projecte

En el directori on hi ha l'aplicació instal·lada existeix una carpeta anomenada 'projects'. En aquest directori, és on s'emmagatzemaran els nous projectes creats que s'emmagatzemin a disc. Cada projecte representa un nou directori que duu el nom del projecte al qual fa referència. Dins d'aquest nou directori, es crea també una carpeta que rep el nom de 'config' i és on s'hi desa la configuració del projecte. Per a tenir més clar l'estructura de directoris, a continuació se'n mostra un exemple:

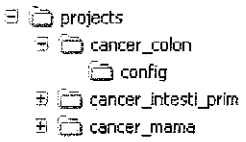


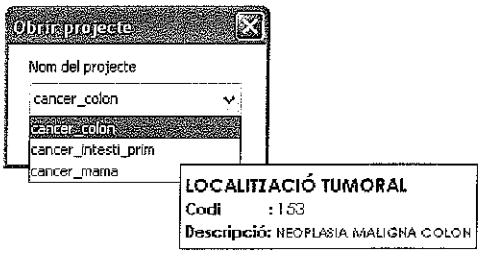
Figura B.3: Estructura del directori 'projects' i els seus subdirectoris

B.3.2 Obrir projecte

En el cas que existeixi algun projecte creat i desitgi obrir-lo, seleccioni l'opció del menú 'Projecte', i després, seleccioni 'Obrir projecte', o bé, premi la segona opció de la barra d'eines de gestió de projectes.

Després de prémer-la, l'aplicació mostrarà una finestra on se li demana que seleccioni el nom del projecte que vol obrir. Tan sols cal que faci un clic damunt d'un nom de projecte i, acte seguit, premi el botó "Accepta". De la mateixa manera que en l'anterior procés, si deixa el cursor damunt de qualsevol nom se li mostrarà la descripció d'aquell projecte.

(a) Seleccionar el nom de projecte



(b) Descripció del projecte

Figura B.4: Procés per a obrir un projecte

B.3.3 Eliminar projecte

De la mateixa manera que en l'anterior funcionalitat, és necessari que existeix un projecte per tal de poder-lo eliminar. Així doncs, tan sols cal que premi l'opció del menú 'Projecte', i dins del desplegable, l'opció 'Eliminar projecte'. També pot accedir-hi prement la tercera opció de la barra d'eines de gestió de projectes.

Un cop hagi seleccionat l'opció d'eliminar un projecte, se li mostrarà una finestra on l'aplicació li demana que esculli el nom del projecte que desitgi eliminar. El fet d'eliminar un projecte comporta esborrar tota la informació que té associada, a més a més, d'eliminar tots aquells fitxers generats a petició de l'usuari en cada consulta executada en aquell projecte en concret. Tan sols cal que seleccioni aquell projecte que vulgui eliminar de la llista de projectes mostrada pel sistema, i acte seguit, premi el botó 'Accepta'.

(a) Seleccionar el nom de projecte

Figura B.5: Procés per a eliminar un projecte

B.3.4 Desar resultats

En cada consulta executada, l'aplicació li demana que seleccioni els fitxers addicionals que vol generar amb els resultats obtinguts en la consulta executada. Si vol emmagatzemar a disc aquests arxius només és necessari que premi l'opció del menú 'Projecte', i a continuació, l'opció 'Desar resultats'. De la mateixa manera, es pot aconseguir el mateix resultat prement la quarta opció de la barra d'eines de gestió de projectes.

Cal recordar com s'estructuren els fitxers dels projectes emmagatzemats a disc (vegeu la figura B.3.1). Tots aquells fitxers generats durant l'execució de la consulta queden organitzats en carpetes, on cadascuna d'ella, representa el conjunt de fitxers emmagatzemats d'un tipus en concret per aquell projecte. Cadascun d'aquests fitxers doncs, rep un nom que permet identificar-ne la consulta a la qual pertanyen i la data de creació. A continuació, se'n mostra un exemple:

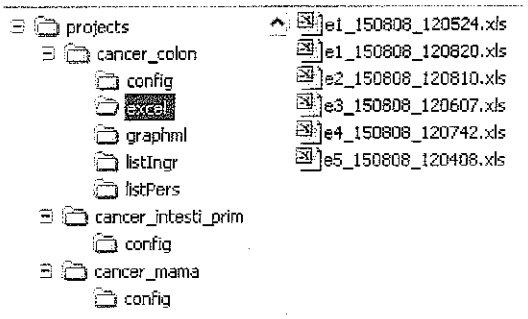


Figura B.6: Estructura del directori 'projects' i els seus subdirectoris i fitxers

B.3.5 Tancar projecte

Per tal de tancar el projecte que s'hagi obert prèviament, només cal que premi l'opció del menú 'Projecte', i dins del desplegable, seleccioni 'Tancar projecte'. També s'aconsegueix els mateixos efectes si es prem el cinquè botó de la barra d'eines de gestió de projectes.

B.4 Execució de consultes

Després d'haver creat o obert un projecte, és el moment d'executar aquelles consultes que li permetin extreure la informació necessària per a dur a terme una anàlisi de la base de dades. El sistema li ofereix un total de cinc consultes partint de la localització tumoral assignada al projecte. A continuació, es pot trobar la definició per a cadascuna d'elles:

Consulta #1 Obté el conjunt de tots els procediments principals i secundaris d'entre totes les altes hospitalàries relacionades amb els individus que se'ls hi va localitzar un tumor determinat en algun moment de la seva vida.

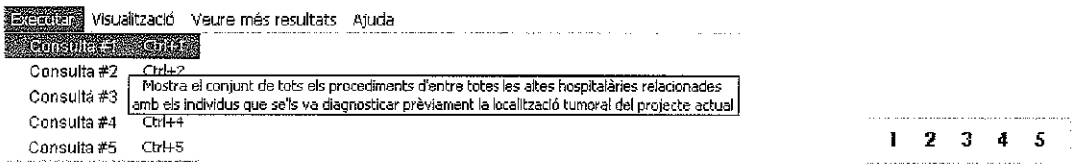
Consulta #2 Obté el conjunt de tots els diagnòstics principals i secundaris d'entre totes les altes hospitalàries relacionades amb els individus que se'ls hi va localitzar un tumor determinat en algun moment de la seva vida.

Consulta #3 Obté el conjunt de tots els procediments principals d'entre totes les altes hospitalàries relacionades amb els individus que se'ls hi va localitzar un tumor determinat en algun moment de la seva vida, classificats pel número d'alta hospitalària i ordenats cronològicament.

Consulta #4 Obté el conjunt dels diagnòstics principals i procediments principals que corresponen a la primera alta hospitalària dels individus que se'ls hi va localitzar un tumor determinat en algun moment de la seva vida.

Consulta #5 Obté el conjunt de tots els procediments principals i diagnòstics principals d'entre totes les altes hospitalàries relacionades amb els individus que se'ls hi va localitzar un tumor determinat en algun moment de la seva vida, classificats pel número d'alta hospitalària i ordenats cronològicament.

El conjunt de consultes mencionades es troben agrupades en el menú 'Executar', o bé, a la barra d'eines d'execució de consultes.



(a) Menú 'Executar' (b) Barra d'eines d'execució de consultes

Figura B.7: Opcions d'execució de consultes

B.4.1 Execució de la consulta N

En cada consulta, els passos a seguir per obtenir informació són els mateixos. Així doncs, la lletra 'N' es refereix a la consulta 1, 2, 3, 4 o 5.

Si desitja executar qualsevol consulta és necessari prémer l'opció del menú 'Executar', i posteriorment, l'opció 'Consulta #N'. També és possible arribar al mateix punt mitjançant el botó amb el número 'N' de la barra d'eines d'execució de consultes.

Després de fer clic en aquesta opció, el sistema li mostra una finestra on apareix una llista multi opció de possibles fitxers addicionals a generar. Aquests fitxers doncs, són el conjunt de resultats representats en diferents formats que s'obtenen en l'execució d'una consulta. A continuació, s'ofereix la descripció de cadascuna d'aquestes opcions:

Generar report	Genera un informe en format textual amb els resultats obtinguts en l'execució de la consulta
Generar excel	Genera un fitxer de Microsfot Excel organitzat en taules amb els resultats obtinguts en l'execució de la consulta
Generar graphML	Genera un fitxer en format graphML que representa, en una estructura de graf escrita en XML, la informació resultant de l'execució d'una consulta
Generar llista persones	Genera un fitxer en format textual que conté la relació entre cada identificador de node de la visualització del graf resultant i la llista de persones que hi estan relacionades
Generar llista d'ingressos	Genera un fitxer en format textual que conté la relació entre cada identificador de node de la visualització del graf resultant i la llista d'ingressos que hi estan relacionats

Cal que seleccioni aquelles opcions que cregui pertinents, i acte seguit, premi el botó 'Accepta'. Després d'haver seguit aquests passos, haurà d'esperar uns instants mentre s'executa la consulta, es generen els resultats en una estructura de graf, es crea la visualització d'aquest i, finalment, es generen els fitxers addicionals que hagi seleccionat prèviament. Un cop finalitzat aquest llarg procés, el sistema li mostrarà els resultats en forma de graf dins de l'aplicació.

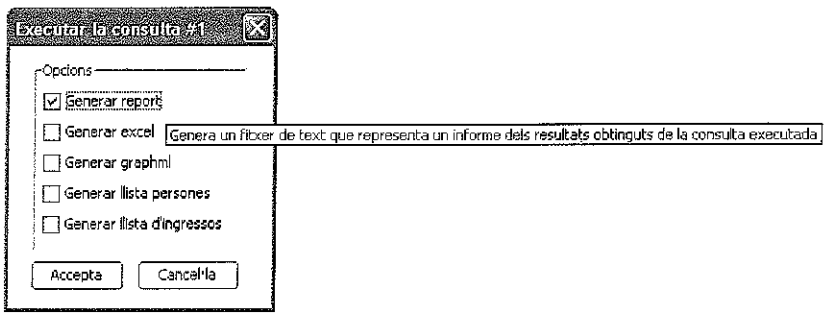


Figura B.8: Procés per a executar una consulta

Cal dir que aquests fitxers addicionals generats s'emmagatzemen en memòria temporal, i més endavant, es detallarà com s'ha de fer per a visualitzar-ne la informació. En el cas que desitgi desar-ho a disc, pot utilitzar la funcionalitat 'Desar resultats' explicada anteriorment.

B.4.2 Visualització del graf resultant

A continuació es defineix quina informació dóna com a resultat cada consulta i com es veu representada en forma de graf en la visualització. A més, per a simplificar-ne l'explicació, s'utilitzarà un exemple basat en la localització tumoral *152 (càncer d'intestí prim)*.

Cal dir que, en cada node, el número que apareix entre claus representa l'identificador de node amb l'objectiu de permetre'n l'accés a algunes dades en concret quan s'escaigui.

Finalment, la visualització del graf resultant està simplificada en tots els exemples per tal de fer-ho més clar i entenedor ja que tan sols apareixerà aquella informació que superi un percentatge mínim de

repeticions per a garantir-ne la rellevància de manera que els percentatges no sumin el 100% com s'esdevé en els exemples reals.

B.4.2.1 Consulta 1

Com es pot veure en la següent figura, existeix un node central que representa el conjunt d'incidències tumorals que coincideixen amb la localització tumoral definida com a punt de partida.

Podem veure com existeixen 392 procediments principals i secundaris aplicats en totes les altes hospitalàries de tots els individus representats en el node arrel. D'aquest node, en surten 7 arestes que contenen la quantitat d'altes hospitalàries i el percentatge equivalent que estan representades i agrupades en el node del seu extrem.

D'aquesta manera, en 83 altes hospitalàries (21,17%) no hi ha hagut cap procediment secundari a aplicar i, en canvi, el procediment "88.01: T.A.C. ABDOMEN" s'ha aplicat un total de 23 vegades (5,86%).

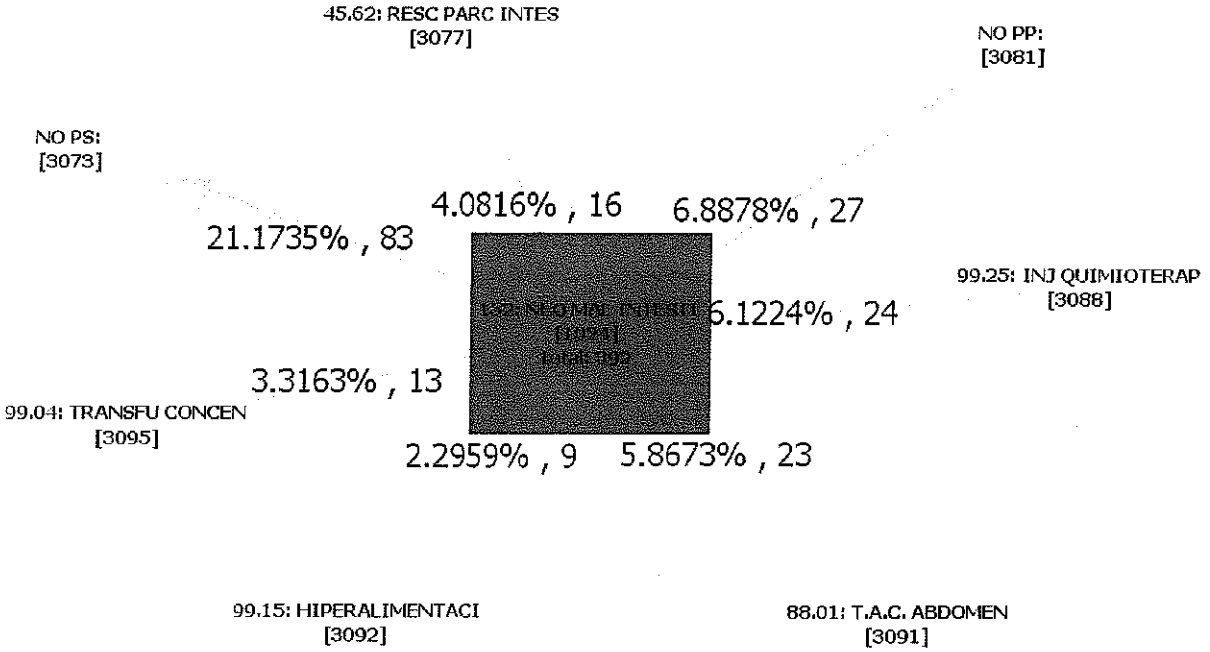


Figura B.9: Exemple de la visualització del graf resultant per la consulta 1

B.4.2.2 Consulta 2

Aquesta consulta té un gran símil amb l'anterior tenint com a node arrel aquell que representa el conjunt d'incidències tumorals de totes aquelles persones que se'ls hi ha localitzat el tumor definit.

En aquest cas, existeixen un total de fins a 149 diagnòstics principals i/o secundaris d'entre totes les altes hospitalàries de totes les persones que es troben representades en el node arrel.

En el següent exemple, podem veure com el node que agrupa a 149 persones i que representa la localització tumoral està unit amb la resta de nodes que representen el conjunt de diagnòstics, tant principals com secundaris. Així doncs, els valors d'aquestes arestes indiquen la quantitat de vegades que es repeteix aquell diagnòstic del node on apunta, i a més, quin percentatge representa entre els nodes del mateix nivell.

En la imatge, per exemple, el diagnòstic que més vegades es repeteix és "152.0: NEO MALIGNA DUO", la qual cosa és raonable ja que coincideix amb el tumor del qual es parteix.

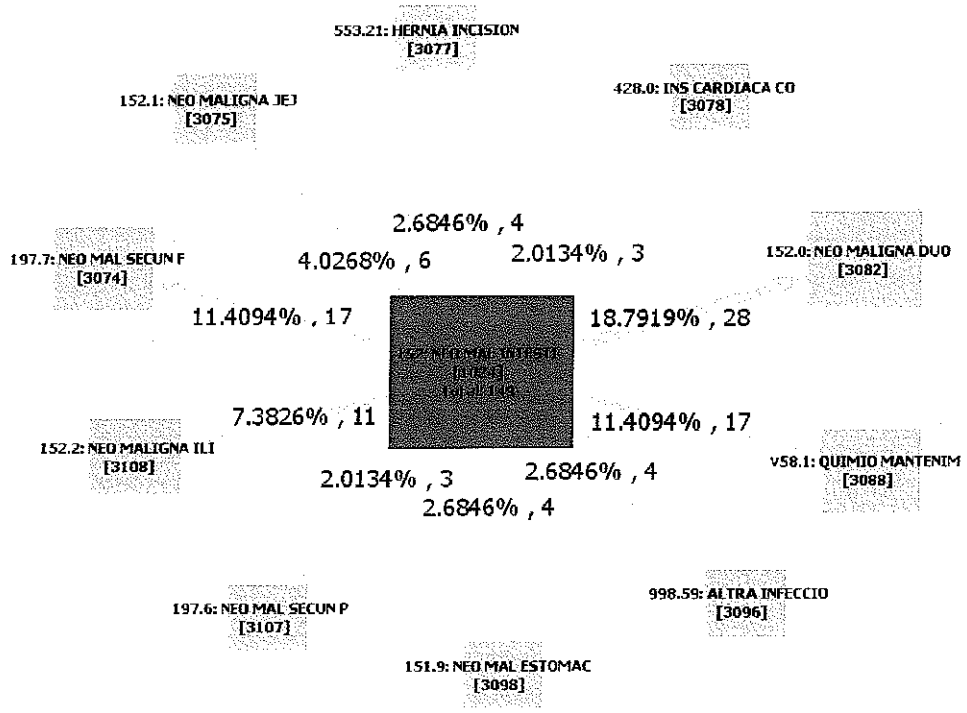


Figura B.10: Exemple de la visualització del graf resultant per la consulta 2

B.4.2.3 Consulta 3

Tot i que la visualització del graf no és complerta ja que s'ha decidit talla la imatge a partir de la tercera alta hospitalària, és suficient per a descriure'n el funcionament.

Partint del node de color vermell que representa el conjunt d'incidències tumorals de totes aquelles persones que se'ls hi ha localitzat el tumor definit, podem veure com existeix una cronologia d'altes hospitalàries.

Del node principal en surten dues arestes que es dirigeixen a dos nodes de color negre que representen el conjunt de persones que han estat ingressades alguna vegada després d'haver-los-hi localitzat el tumor. L'altre node de color negre, representa aquells que no han estat ingressats cap vegada després de la localització tumoral.

En aquest exemple doncs, 62 persones van ingressar almenys una vegada i 66 no ho van fer mai. D'aquestes 62, també se n'ofereix representada en el mateix node, la mitjana de dies que van transcórrer entre la data d'incidència del tumor i el seu primer ingrés. Aquesta xifra és de 155,871 dies.

Del node que representa les 62 persones que tenen registrada almenys una alta hospitalària en surt una altra aresta que apunta a un node ataronjat. Aquest node doncs, representa un procediment principal mitjançant un codi de dos dígits.

En aquest cas, només n'apareix un per a simplificar-ho però, en realitat, n'hi ha més com es pot deduir amb el percentatge de l'aresta.

Cal afegir que ara, l'aresta conté dos percentatges. El primer indica el percentatge del node respecte el conjunt de persones indicades en el node vermell, i el segon, indica el percentatge del node respecte els nodes del seu mateix nivell.

Així doncs, un subconjunt de 7 persones del total de 62 persones esmentades anteriorment se'ls hi va aplicar el procediment "99: ALT TECNIQUES" en la seva primera alta hospitalària.

D'aquest node ataronjat, en surten 4 arestes de color gris i 2 de color taronja. Les dues arestes de color taronja apunten a 2 nodes de color negre. El primer node representa tots aquells individus que ja no van reingressar cap més vegada, i el segon, el total de persones que van tornar a ser ingressades per segona vegada. Les 4 arestes grises apunten a 4 nodes de color groc. Aquestos nodes són, de la mateixa manera que el seu node pare, procediments principals amb la diferència que aquestos contenen 4 dígits. Així doncs, si el node pare representava un procediment principal aplicat a un total de 7 persones, aquestos 4 nodes, representen el detall dels procediments principals aplicats a aquestes 7 persones.

A partir d'aquí, la descripció torna a ser la mateixa que la descrita a partir del node que representa la primera alta hospitalària.

Cal mencionar també, que els continguts de les arestes de tot el graf representen el mateix en cada moment.

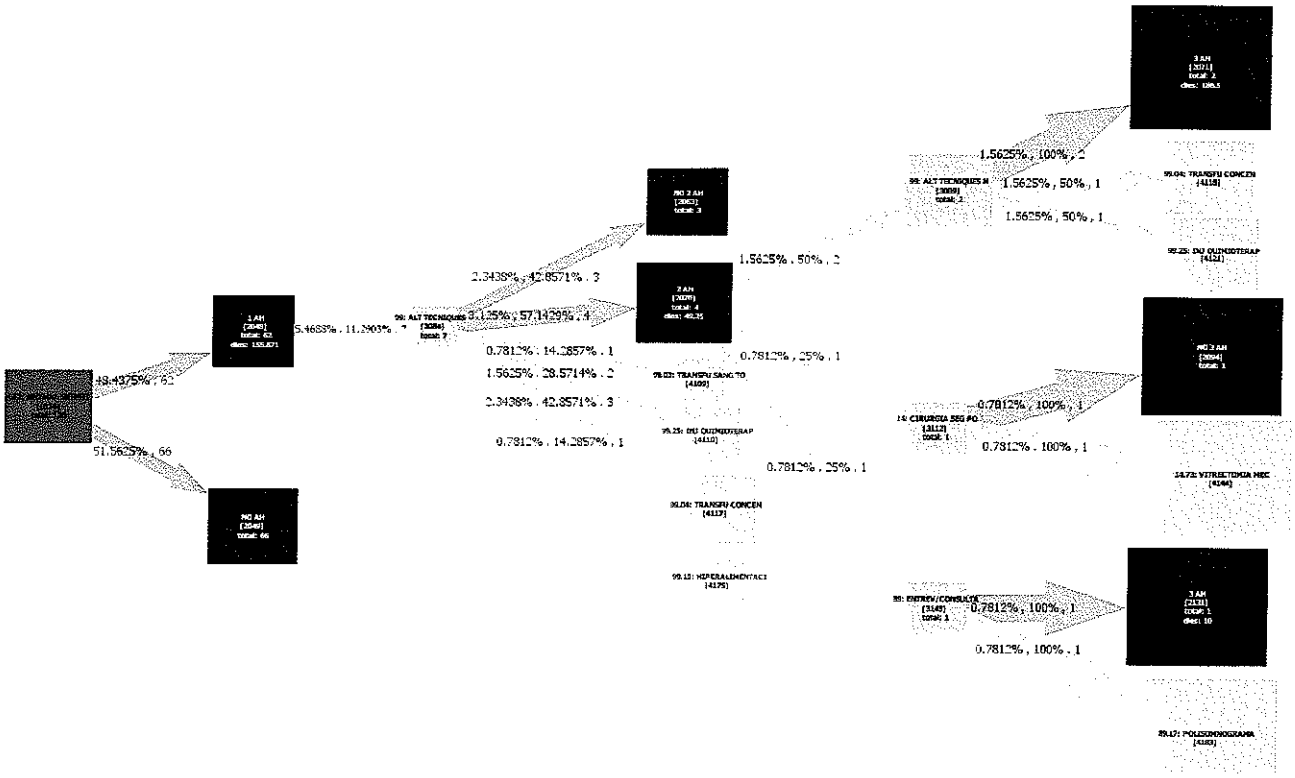


Figura B.11: Exemple de la visualització del graf resultant per la consulta 3

B.4.2.4 Consulta 4

Aquesta consulta té un comportament similar a la descrita anteriorment amb la diferència que tan sols es centra en la representació de tota la informació relacionada amb el primer ingrés.

Així doncs, el node del qual parteix tot continua sent el conjunt d'incidències d'un tumor en concret. De la mateixa manera que en l'anterior, se'ls hi va localitzar el tumor definit a un total de 128 persones.

D'aquest total d'individus en surten dues arestes. La primera es dirigeix cap al node de color negre que representa el conjunt de persones que van ingressar almenys una vegada que, òbviament, continua sent de 62 persones, i la segona, apunta al node que agrupa al total d'individus sense cap alta hospitalària posterior a la data d'incidència.

Trencant amb la descripció de la consulta anterior, les dues arestes que continuen del node que representa la segona alta hospitalària apunten a dos nodes de color taronja que es defineixen com a diagnòstics principals. Així doncs, 29 persones del total de 62 persones ingressades van ser diagnosticades de "152: NEO MAL INTESTI", i 10 ho van ser de "197: NEO MAL SECUN D".

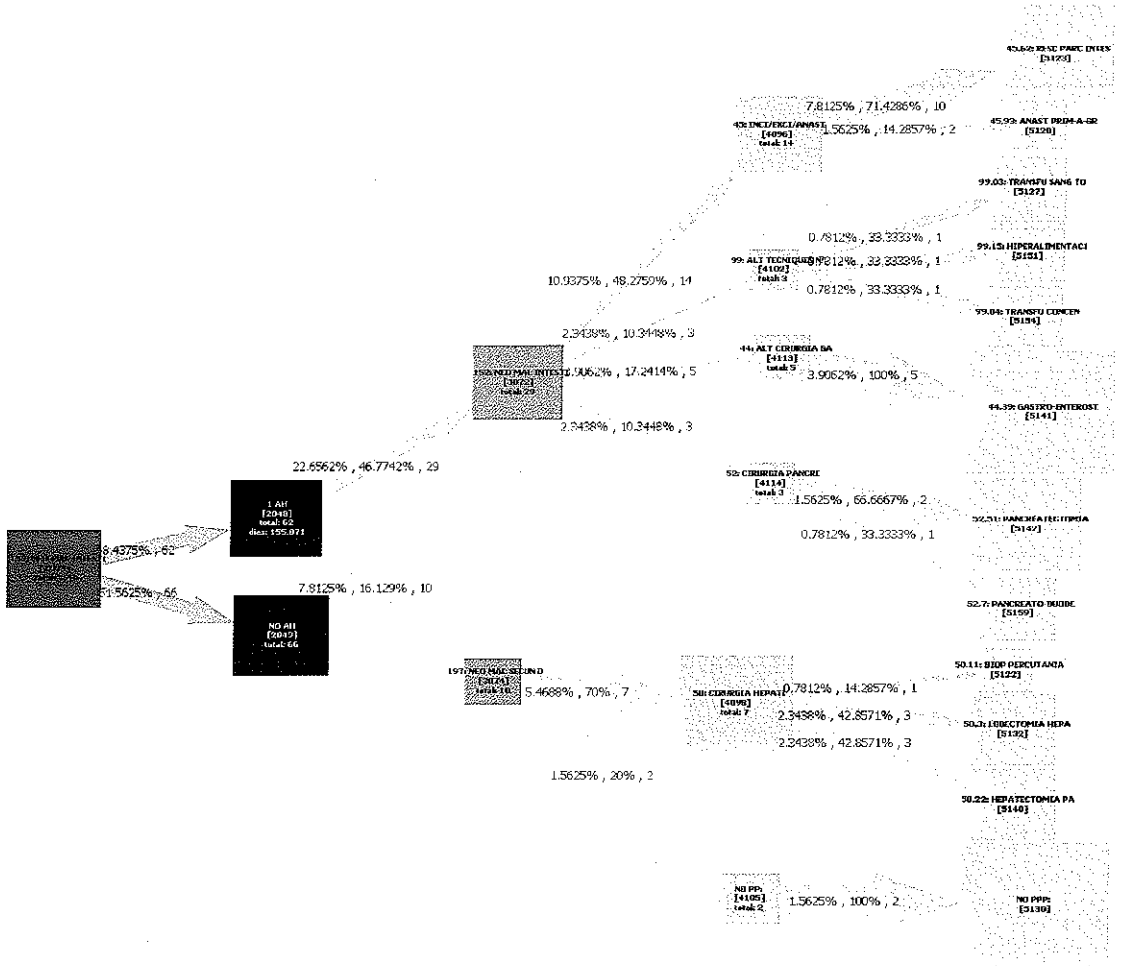


Figura B.12: Exemple de la visualització del graf resultant per la consulta 4

A partir d'aquests 2 nodes, comencen a sortir més arestes que apunten altre cop a nodes de color ataronjat. Aquests nodes, representen els procediments principals aplicats en la primera alta hospitalària que comparteixen el mateix diagnòstic principal representat en el seu node pare. D'aquesta manera, del total de 10 persones amb el diagnòstic principal de "197: NEO MAL SECUN D", a 7 se'ls hi va aplicar un procediment principal de "50: CIRURGIA HEPAT" i a les altres 3 no se'ls n'hi va aplicar cap.

Finalment, de cada node ataronjat surten vàries arestes que es dirigeixen cap a nodes de color groc. Aquests, representen el conjunt de persones que se'ls hi ha aplicat el procediment principal que indica el mateix node d'igual manera que ho feia el pare. L'únic diferència rau en què els nodes fills indiquen amb més detall quins són aquests procediments principals gràcies als seus 4 dígits enlloc dels 2 dígits.

Així doncs, de les 7 persones que se'ls hi va aplicar el procediment principal "50: CIRURGIA HEPAT", a 3 se'ls hi va aplicar concretament el procediment "50.3: LOBECTOMIA HEPA", a uns altres 3 va ser "50.22: HEPATECTOMIA PA" i a l'altra persona que resta va ser-li aplicat un procediment de "50.11: BIOP PERCUTANIA".

B.4.2.5 Consulta 5

La descripció del resultat de l'execució de la consulta anterior es totalment aplicable a aquesta consulta. Tan sols cal afegir que, a diferència de la consulta 4, aquesta permet obtenir la mateixa informació per a totes les altes hospitalàries i no tan sols per la primera.

Concretament, permet visualitzar la cronologia de la informació definida en la consulta anterior, per la qual cosa no és necessari expandir-se en la seva explicació ja que s'ha definit el mateix prèviament en la consulta 4.

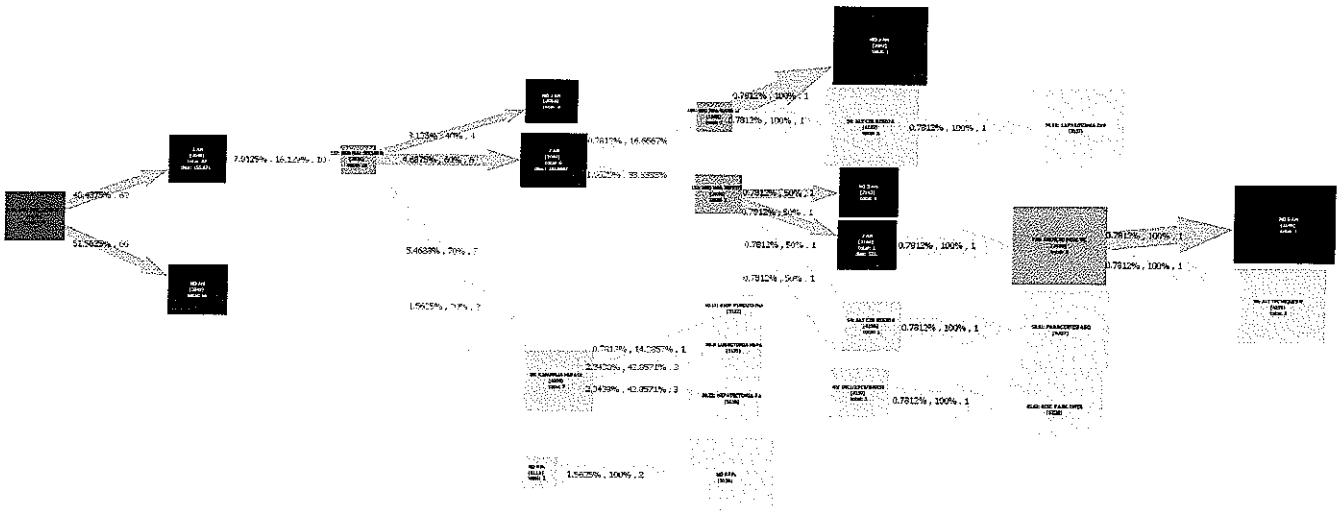


Figura B.13: Exemple de la visualització del graf resultant per la consulta 5

B.5 Llegenda de la visualització del graf

Per tal d'entendre l'estructura de la visualització del graf resultant de la consulta, a continuació s'explicarà amb un exemple, que representa cadascun dels elements que hi apareixen:

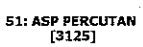
B.6 Visualització del graf

Un cop es disposa de la visualització del graf resultant que representa la informació obtinguda en l'execució de la consulta, es pot modificar algunes opcions amb l'objectiu d'obtenir més informació o resumir-la, a més a més, de canviar algunes distàncies entre nodes. A continuació, es descriu quina funció desenvolupa cadascuna d'aquestes opcions:

Conjunt d'individus que se'ls hi ha localitzat el tumor amb una descripció curta NEO MAL INTESTI i codi 152. L'identificador del node és el 1024, i finalment, formen aquest grup 392 persones.

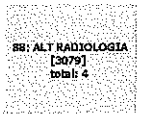
Conjunt d'individus que se'ls hi ha registrat la primera alta hospitalària després d'haver-los-hi localitzat un tumor. L'identificador de node és el 2048, i el total de persones que formen el grup és de 62. La mitjana de dies, entre tots els individus, que han passat entre aquesta alta hospitalària i l'anterior és de 155,871 dies.

Conjunt d'altes hospitalàries en què s'ha registrat un diagnòstic principal amb la descripció NEO MAL ESTOMAC i codi 151.9. En el codi, el número que apareix a la dreta del punt indica l'especificació del diagnòstic genèric amb codi 151. L'identificador del node és el 3098.

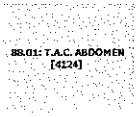


Conjunt d'altes hospitalàries en què s'ha registrat un procediment principal i/o secundari amb la descripció ASP PERCUTAN i codi 51. L'identificador del node és el 3125.

Conjunt d'individus que se'ls hi ha registrat un diagnòstic principal amb la descripció NEO MAL INTESTI i codi 152 en un número d'alta hospitalària determinat. L'identificador del node és el 3072, i el total de persones que formen el grup és de 29.



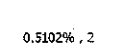
Conjunt d'individus que se'ls hi ha registrat un procediment principal amb la descripció ALT RADIOLOGIA i codi 88 en un número d'alta hospitalària determinat. L'identificador del node és el 3079, i el total de persones que formen el grup és de 4.



Conjunt d'individus que se'ls hi ha registrat un procediment principal amb la descripció T.A.C. ABDOMEN i codi 88 en un número d'alta hospitalària determinat. El número que apareix a la dreta del punt indica l'especificació del procediment primari amb codi 88. L'identificador del node és el 4124.

Aresta que representa la relació entre un diagnòstic principal o localització tumoral i una alta hospitalària.

Aresta que representa la relació entre la resta de nodes no mencionats en la descripció anterior.



Etiqueta d'aresta en la qual el percentatge indica la representació del node al qual apunta l'aresta respecte els nodes del seu mateix nivell, i el nombre de la dreta, indica la quantitat d'individus agrupats en aquest node mencionat.

Etiqueta d'aresta en la qual el primer percentatge indica la representació del node al qual apunta l'aresta respecte el conjunt d'individus agrupats en el node arrel, el segon percentatge indica la representacio respecte la resta de nodes del seu mateix nivell, i l'últim nombre indica la quantitat d'individus agrupats en aquest node mencionat.

Visualitzar el graf sencer	Canvia la visualització del graf per a què es mostri el graf complert, és a dir, tots els elements que conformen el conjunt de la informació obtinguda en la consulta
Visualitzar el graf resum	Canvia la visualització del graf per a què es mostri un resum del graf resultant, és a dir, agrupant alguns nodes en un de genèric
Modificar distància amplada	Augmentar o disminuir la distància entre els nodes del mateix nivell
Modificar distància profunditat	Augmentar o disminuir la distància entre els nodes de diferent nivell
Opcions de filtratge	Canviar la visualització de manera que tan sols es mostrin aquells nodes del graf resultant que superin el valor de filtratge introduït que es refereix al seu percentatge respecte els nodes del mateix nivell, o a la quantitat de vegades que es repeteix la informació del node.

Aquestes funcionalitats estan situades al menú 'Visualització', o bé, a la barra d'eines d'opcions de visualització.

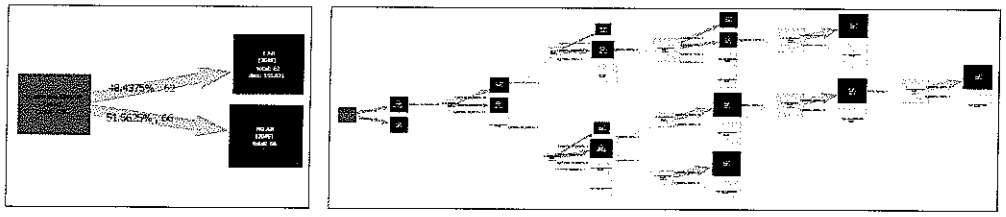
(a) Menú 'Visualització' (b) Barra d'eines d'opcions de visualització

Figura B.14: Opcions de visualització

B.6.1 Visualitzar el graf sencer

Si es vol visualitzar el graf complert, només cal seleccionar l'opció del menú 'Visualització', i acte seguit, 'Graf Complert'. A més a més, pot accedir a la mateixa opció mitjançant el primer botó de la barra d'eines d'opcions de visualització.

El sistema esborrarà l'actual visualització i hi mostrarà la nova.

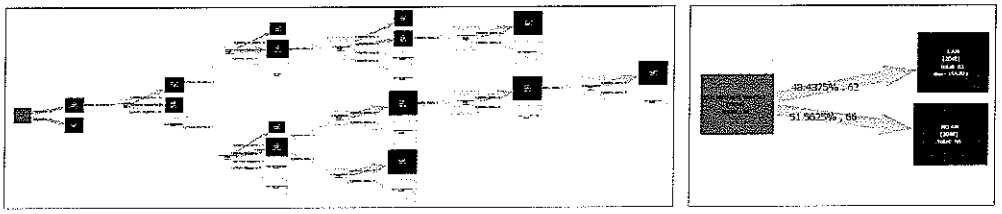


(a) Moment inicial (b) Moment final

Figura B.15: Exemple de visualització del graf sencer

B.6.2 Visualitzar el graf resum

Per tal de canviar la visualització del graf sencer a un graf resum premi l'opció del menú 'Visualització'. El sistema li mostrarà un desplegable i ha de seleccionar l'opció 'Graf Resum'. També és possible visualitzar el graf resum mitjançant el segon botó de la barra d'eines d'opcions de visualització.



(a) Moment inicial (b) Moment final

Figura B.16: Exemple de visualització del graf resum

B.6.3 Modificar distància amplada

La distància que existeix entre els nodes del mateix nivell pot ser modificada mitjançant l'opció del menú 'Visualització', i a continuació, 'Opcions de visualització'. És possible arribar al mateix punt tan sols prement el tercer botó de la barra d'eines d'opcions de visualització.

En prémer aquest botó, si la consulta executada és la tercera, la quarta o la cinquena, el sistema li mostrarà una finestra on hi apareixen dues barres de desplaçament.

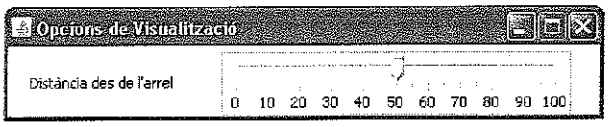
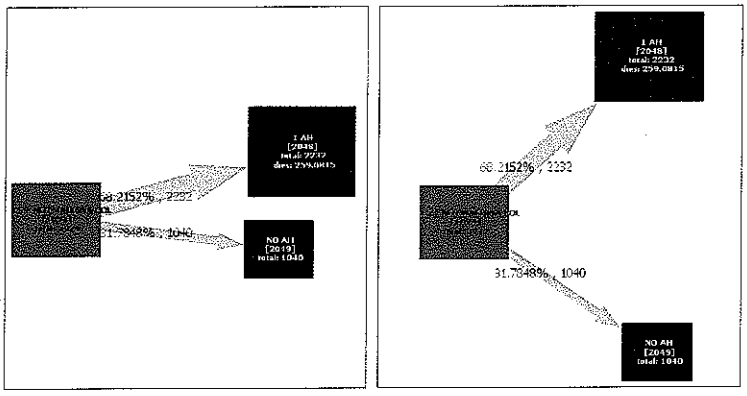


Figura B.17: Opcions de modificació de distància en amplada

Per tal de canviar aquesta distància és necessari moure la segona barra anomenada 'Espai en amplada'. Si es vol augmentar aquesta distància cal que desplaci la barra cap a la dreta, però si al contrari vol reduir-la, haurà de moure-la cap a l'esquerra.



(a) Moment inicial (b) Moment final

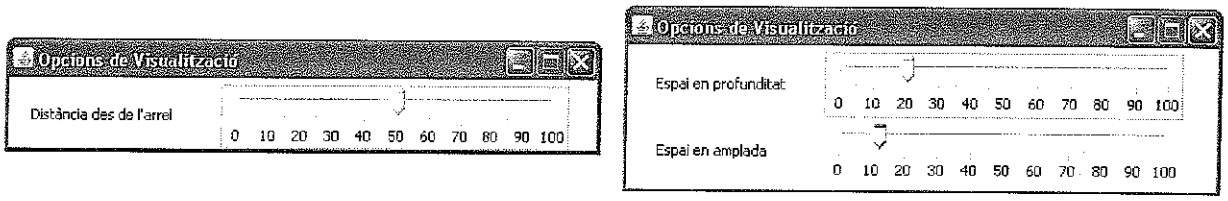
Figura B.18: Exemple d'augment de la distància en amplada

B.6.4 Modificar distància profunditat

Per tal de modificar la distància que existeix entre els nodes de diferent nivell, és necessari prémer l'opció del menú 'Visualització', i a continuació, 'Opcions de visualització'. També ho pot fer mitjançant el tercer

botó de la barra d'eines d'opcions de visualització.

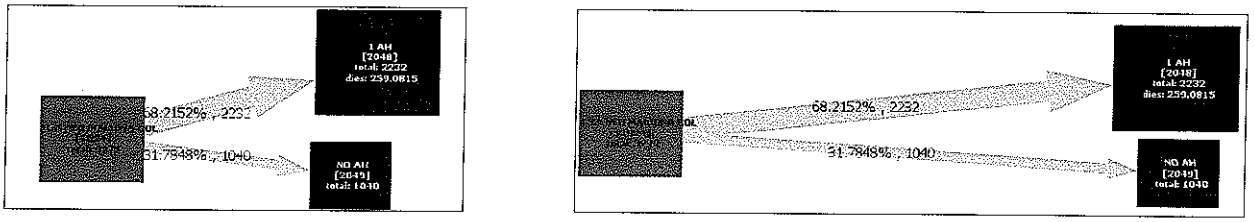
Aleshores, depenent de la consulta executada, apareixeran una o dues barres de desplaçament. Si aquesta consulta és la primera o la segona, tan sols es mostrarà una barra amb el nom 'Distància des de l'arrel'. Si en canvi, la consulta és la tercera, quarta o cinquena, apareixeran dues barres.



(a) Consulta 1 i 2 (b) Consulta 3, 4 i 5

Figura B.19: Opcions de modificació de distància en profunditat

Si es desitja canviar aquesta distància tan sols cal desplaçar les barres mencionades anteriorment. De la mateixa manera que en l'anterior funcionalitat, si es desplaça la barra cap a la dreta, s'aconsegueix que augmenti la distància, i si es fa a la inversa, es veu reduïda.



(a) Moment inicial (b) Moment final

Figura B.20: Exemple d'augment de la distància en profunditat

B.6.5 Opcions de filtratge

Si es vol modificar la visualització per tal de mostrar només els nodes que superen el valor de filtratge, tan sols cal seleccionar l'opció del menú 'Visualització', i dins del desplegable, el botó 'Opcions de filtratge', o bé, prement el quart botó de la barra d'eines d'opcions de visualització.

A continuació, apareixerà una finestra on s'hi poden veure dues opcions i un requadre. Si es desitja aplicar un filtratge pel número de repeticions cal que seleccioni l'opció 'Repeticions'. Si en canvi, vol filtrar la visualització pels percentatges dels nodes, serà necessari que elegeixi l'opció 'Percentatge'. Un cop escollida l'opció, ha d'introduir el valor de filtratge dins del requadre.

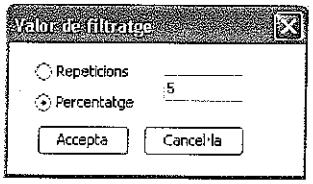
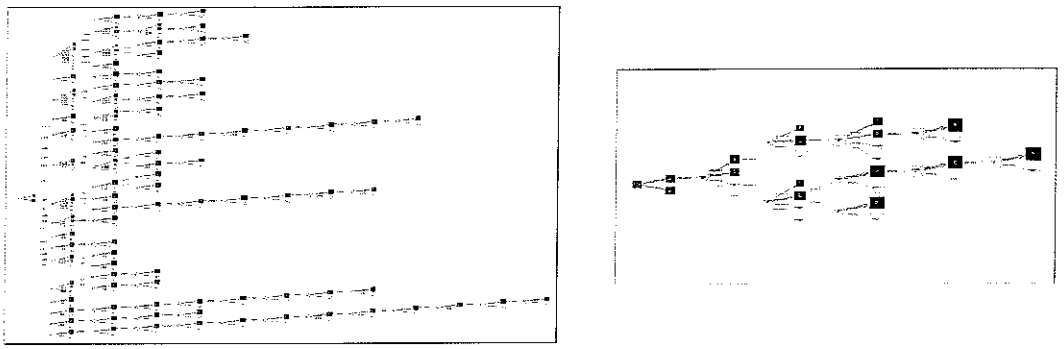


Figura B.21: Opcions de filtratge

Finalment, si es vol aplicar el filtre a la visualització, tan sols cal prémer el botó 'Accepta', i a continuació, s'esborrarà l'actual visualització i se'n generarà una de nova amb els canvis introduïts.



(a) Moment inicial (b) Moment final

Figura B.22: Exemple de filtratge per percentatge

B.7 Manipulació de la visualització del graf

Després d'haver indicat al sistema quina consulta ha de ser executada i que, aquest, hagi dut a terme els passos necessaris per a la creació i visualització dels resultats, l'usuari pot dur a terme accions sobre la visualització per tal de millorar-la i adaptar-la a les seves necessitats.

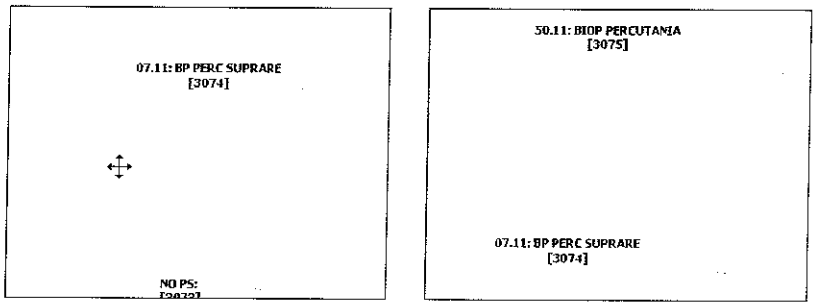
Així doncs, les diferents opcions de manipulació del graf són les següents:

Desplaçar àrea visualització	Permet desplaçar l'àrea de visualització per tal que es pugui canviar la zona del total del graf a mostrar
Modificar àrea visualització	Permet augmentar o disminuir l'àrea de visualització amb l'objectiu de veure més grans els elements del graf però veure'n menys, o a la inversa, veure més elements del graf però veure'ls més petits
Encabir el graf	Permet encabir el graf sencer dins la visualització, mostrant així tots els elements dels quals està compost
Seleccionar àrea visualització	Permet seleccionar, amb un requadre, quina ha de ser la nova àrea de visualització
Expandir node	Permet expandir tots els nodes que pengen del node seleccionat
Mostrar informació node	Permet obtenir la informació detallada que fa referència al node seleccionat

El conjunt de manipulacions descrites es duen a terme, única i exclusivament, amb el ratolí.

B.7.1 Desplaçar àrea visualització

Per tal de desplaçar-ne l'àrea, tan sols és necessari clicar amb el botó esquerre del ratolí damunt de la visualització i arrossegar-ne l'àrea fins que se n'aconsegueixi la posició idònia. En el següent exemple, podem veure com l'àrea de visualització s'ha desplaçat cap amunt:

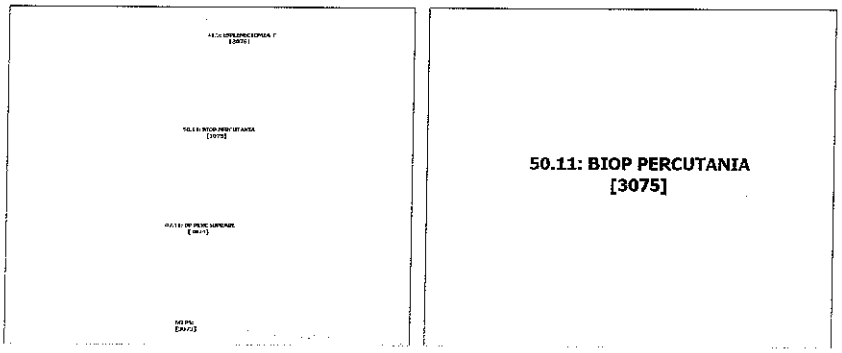


(a) Moment inicial (b) Moment final

Figura B.23: Exemple de desplaçament de l'àrea de visualització

B.7.2 Modificar àrea visualització

Es pot modificar l'àrea de visualització utilitzant la roda (*scroll*) del ratolí. Així doncs, si es fa girar la roda cap amunt augmenta l'àrea de visualització. En canvi, si es fa girar la roda cap avall s'aconsegueix reduir l'àrea. A continuació s'ofereix un exemple:

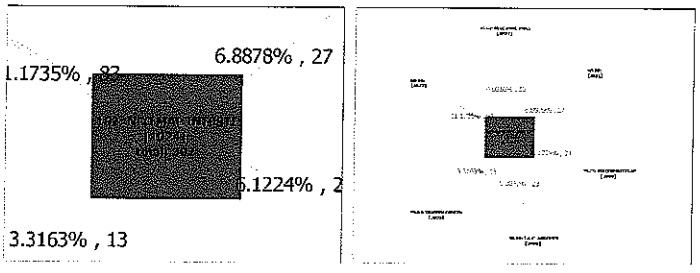


(a) Moment inicial (b) Moment final

Figura B.24: Exemple d'augment de l'àrea de visualització

B.7.3 Encabir el graf

Per tal d'encabir tot el graf dins la visualització tan sols és necessari clicar amb el botó dret del ratolí damunt de la visualització.

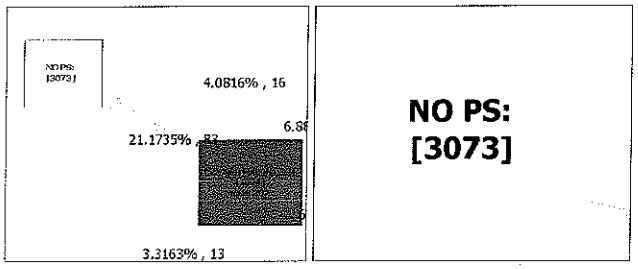


(a) Moment inicial (b) Moment final

Figura B.25: Exemple d'encabir el graf dins la visualització

B.7.4 Seleccionar l'àrea de visualització

Si es vol marcar l'àrea de la visualització que es desitja mostrar només cal clicar amb el botó esquerre del ratolí i arrossegar-lo fins que es formi un requadre que correspondrà a la nova àrea.

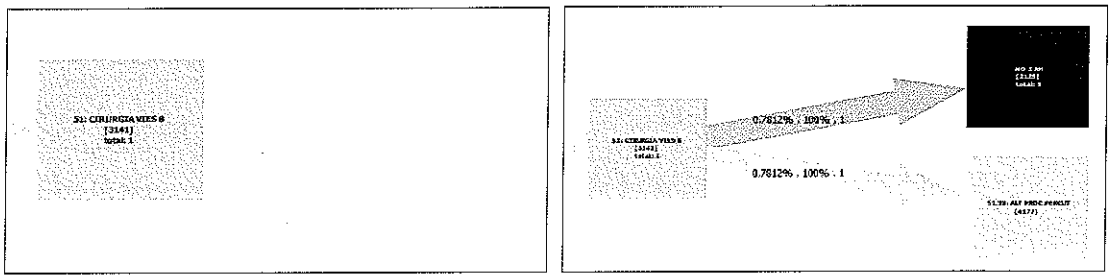


(a) Moment inicial (b) Moment final

Figura B.26: Exemple de selecció de l'àrea de visualització

B.7.5 Expandir node

Per tal de mostrar els nodes que pengen d'un node en concret, és necessari clicar amb el botó esquerre del ratolí damunt del node.



(a) Moment inicial (b) Moment final

Figura B.27: Exemple d'expansió de node

B.7.6 Mostrar informació node

Finalment, si el que es desitja és obtenir informació detallada d'un node, tan sols cal passar el cursor del ratolí per damunt del node en concret i apareixerà un panell amb les dades.

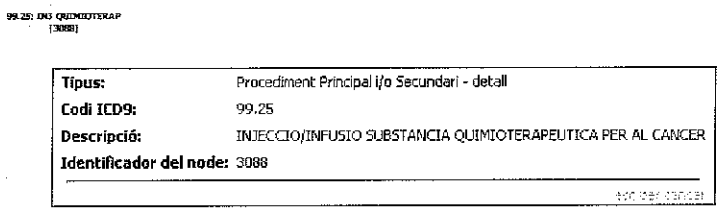


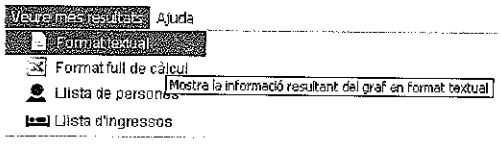
Figura B.28: Exemple de visualització de la informació d'un node

B.8 Veure altres resultats

Després d'executar cada consulta, si s'ha escollit que es generin fitxers addicionals que contenen la informació resultant en altres formats, el sistema ofereix un conjunt de funcionalitats per tal de visualitzar-ne aquells que s'hagin generat prèviament. A continuació, es descriu quines funcions permeten mostrar aquests resultats:

Format textual	Visualitzar la mateixa informació obtinguda en la consulta i representada en la visualització del graf, però estructurada en un fitxer de text organitzat mitjançant l'ús de tabuladors
Format full de càlcul	Visualitzar la mateixa informació obtinguda en la consulta i representada en la visualització del graf, però estructurada en un fitxer de Microsoft Excel organitzat en diferents taules
Cercar persones	Cercar la llista d'identificador de persones relacionada a un identificador de node de la visualització del graf
Cercar ingressos	Cercar la llista d'identificador d'ingressos relacionada a un identificador de node de la visualització del graf

Totes aquestes funcions les pot trobar en el menú 'Veure més resultats', i també, a la barra d'eines de visualització de més resultats.



(a) Menú 'Veure més resultats'

(b) Barra d'eines de visualització de més resultats

Figura B.29: Opcions de visualització

B.8.1 Format textual

Per tal de mostrar la mateixa informació de la visualització del graf en un fitxer de text, tan sols cal que seleccioni l'opció del menú 'Veure més resultats', i acte seguit, premi 'Format textual'.

A continuació, li apareixerà una nova finestra on s'hi representa la citada informació. A més a més, es poden realitzar cerques dins de la visualització dels resultats en format textual. Tan sols cal introduir el text a cercar en el requadre per a escriure situat a la part inferior de la nova finestra i prémer la tecla 'Intro'. El text trobat apareixerà subratllat.

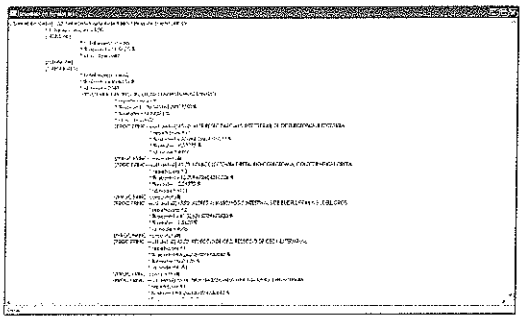


Figura B.30: Resultats en format textual

B.8.2 Format full de càlcul

En aquest cas, per tal de visualitzar la informació resultant de la consulta en format de full de càlcul tan sols cal prémer l'opció del menú 'Veure més resultats', i a continuació, seleccionar l'opció 'Format full de càlcul', o bé, prémer el segon botó de la barra d'eines de visualització de més resultats.

D'aquesta manera, s'executarà l'aplicació de Microsoft Excel i s'obrirà el fitxer generat prèviament que conté la informació mencionada.

B.8.3 Cercar persones

Per tal de cercar la llista dels identificadors de persona relacionats a un identificador de node, cal que seleccioni l'opció del menú 'Veure més resultats', i dins el desplegable, l'opció 'Llista de persones'. També es pot accedir al mateix lloc prement el tercer botó de la barra d'eines de visualització de més resultats.

Després de prémer el botó adient, apareix una nova finestra on s'hi pot veure una taula amb només una fila. La primera columna d'aquesta taula conté l'identificador d'un node de la visualització del graf, i la resta, contenen els diferents identificadors de persona relacionats amb aquell node.

Per defecte, apareix la llista associada a l'identificador del node arrel, però si es desitja veure'n la llista relacionada a un altre identificador de node, només fa falta introduir-lo en el requadre destinat a aquesta funció i prémer la tecla 'Intro'. Les dades de la taula es refrescaran seguint la nova petició.

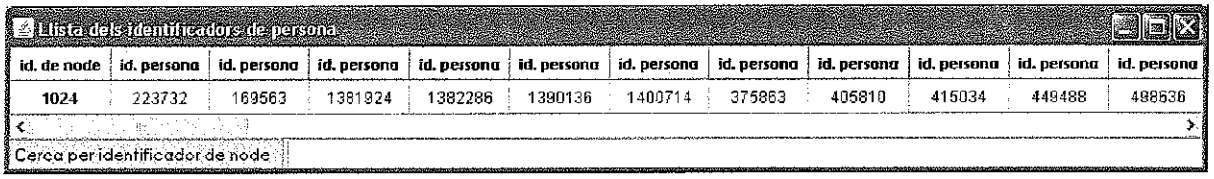


Figura B.31: Exemple de cerca de la llista de persones relacionades a un node

B.8.4 Cercar ingressos

De la mateixa manera que en l'anterior funcionalitat, per tal d'obtenir la llista d'ingressos relacionada a un node en concret, tan sols és necessari seleccionar l'opció del menú 'Veure més resultats', i acte segut, prémer l'opció 'Cercar ingressos'. També es pot arribar al mateix punt prement el quart botó de la barra d'eines de visualització de més resultats.

El procés per a obtenir la llista d'identificadors d'ingressos relacionada a un identificador de node és exactament el mateix que en l'anterior apartat.

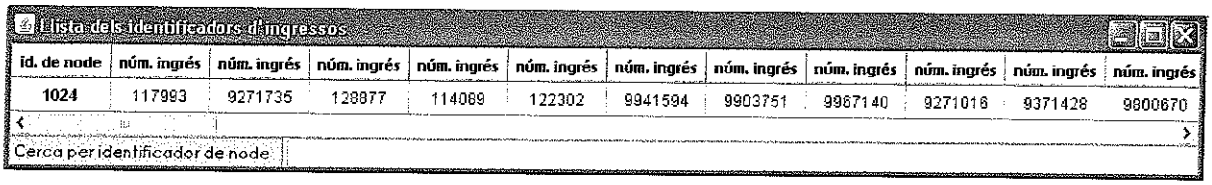


Figura B.32: Exemple de cerca de la llista d'ingressos relacionats a un node

B.9 Cerca de nodes i visualització de l'etiqueta de nodes i arestes

Existeix la possibilitat de realitzar cerques en els nodes de la visualització del graf, així com, obtenir el text contingut en cada node i aresta. A continuació, se'n descriu les dues funcionalitats:

Visualització etiqueta node i aresta — Visualitzar, en una mida de lletra clara i en una posició fixa, el text contingut en l'etiqueta d'un node o aresta

Cerca de nodes — Cercar un node dins la visualització del graf a partir d'un text que ha de coincidir amb alguna part del text contingut en l'etiqueta del node

Aquestes funcions es troben agrupades en un panell de color blau sota la visualització del graf.

Figura B.33: Panell de cerca de nodes i visualització de l'etiqueta

B.9.1 Visualització etiqueta node i aresta

Per tal de veure el contingut de l'etiqueta per a cada node i aresta de la visualització del graf tan sols és necessari passar el cursor damunt de cadascun d'aquests elements i, automàticament, apareixerà l'etiqueta d'aquest en l'espai destinat a aquesta funció. Aquest espai és just aquella part del panell que queda a la dreta del requadre anomenat 'Contingut del node/aresta' tal i com el seu nom indica.

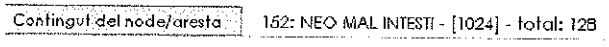
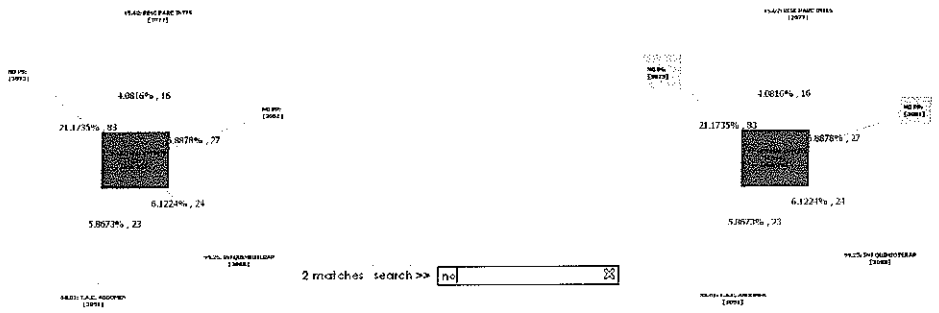


Figura B.34: Exemple de visualització de l'etiqueta d'un node

B.9.2 Cerca de nodes

Si es vol cercar els nodes que la seva etiqueta contingui el text indicat, només s'ha d'introduir el text dins el requadre situat en el panell mencionat. Aquest requadre es troba situat a la dreta de l'etiqueta anomenada 'search', en l'extrem dret del panell. Aleshores, quan l'etiqueta d'alguns nodes contingui el text introduït, aquest conjunt de nodes passen a ser de color blau per tal d'indicar-ne el resultat de la cerca. Cal dir que a mesura que es va escrivint es va cercant en la visualització del graf.



(a) Moment incial (b) Text a cercar i nodes trobats (c) Moment final

Figura B.35: Exemple de cerca de nodes

B.10 Panells informatius

Existeixen tres panells que informen d'algunes accions de l'aplicació, consultes executades o informació que fa referència al projecte obert.

Així doncs, a continuació se'n descriu la funció de cadascun d'ells:

Panell informatiu de tasques	Mostra les diferents tasques que realitza el sistema a mesura que l'usuari realitza interaccions amb l'aplicació
Panell informatiu de consultes	Mostra la informació que fa referència a la consulta executada i, indica si la visualització del graf és sencera o es mostra un graf resum, a part també, d'indicar si s'ha aplicat algun tipus de filtratge.
Panell informatiu de projecte	Mostra la informació que fa referència al nom del projecte obert i la localització tumoral associada

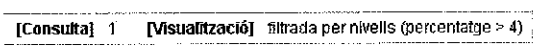
En tots aquests panells no és necessari realitzar cap acció explícitament sinó que el sistema els va actualitzant a mesura que l'usuari va indicant ordres al sistema.

B.10.1 Panell informatiu de tasques

Es troba situat a la part inferior esquerra de l'aplicació.

B.10.2 Panell informatiu de consultes

Es troba situat a la part inferior dreta de l'aplicació.

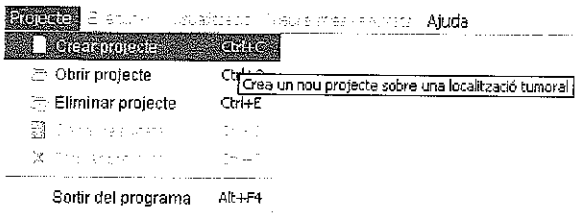


B.10.3 Panell informatiu de projecte

Es troba situat a la part superior dreta de l'aplicació.

B.11 Sortir del programa

Finalment, per sortir del programa, cal prémer l'opció del menú 'Projecte', i acte seguit, 'Sortir del programa'; o bé, prémer el botó implementat per aquesta funció en la part superior dreta.



(a) Menú 'Projecte'

(b) Icona per sortir del programa

Figura B.36: Opció de sortir del programa

Apèndix C

Terminologia

Àrea de visualització Conjunt d'elements del graf resultant que es visualitzen en l'espai destinat per a aquesta funció.

Aresta Línia o segment que connecta dos vèrtexs/nodes d'un graf.

Centre sanitari Institució dedicada a l'atenció, la promoció i l'educació de la salut de la població humana d'un territori en concret.

Diagnòstic Determinació d'una malaltia pels signes o símptomes que li són propis.

Diagnòstic principal Quan existeix més d'un diagnòstic, és la determinació d'aquell que coincideix amb els màxims signes o símptomes que presenta l'individu.

Diagnòstic secundari Quan existeix més d'un diagnòstic, és la determinació d'aquells que no són el diagnòstic principal.

Epidemiòleg Metge especialista en epidemiologia.

Epidemiologia Especialitat científica que estudia la distribució, freqüència, determinants, relacions, prediccions i control dels aspectes relacionats amb la salut i les malalties que afecten a la població humana.

Especificació de la localització tumoral Indica, més detalladament, el tipus de tumor.

Histologia Ciència que estudia el cos humà utilitzant com a instrument el microscopi amb l'objectiu de comprendre el funcionament dels components corporals mitjançant els coneixements anatòmics, bioquímics i fisiològics.

ICD9 És el sistema oficial de codificació de diagnòstics i procediments associats als registres dels centres hospitalaris.

Incidència tumoral Diagnòstic d'un tumor en concret durant la vida d'un individu.

Localització tumoral Indica un tipus de tumor.

Malaltia concomitant Malaltia que acompanya a una altra que n'és l'origen.

Neoplasia maligna Sinònim de càncer.

Node Element del conjunt fonamental d'un graf.

Procediment Conjunt d'accions que es duen a terme amb l'objectiu de determinar, curar o pal·liar una malaltia.

Procediment diagnòstic Conjunt d'accions que es duen a terme amb l'objectiu de determinar una malaltia com a diagnòstic.

Procediment principal (PP) Quan existeix més d'una malaltia, és el conjunt d'accions que es duen a terme amb l'objectiu de determinar, curar o pal·liar aquella que requereixi més atenció.

Procediment secundari (PS) Quan existeix més d'una malaltia, és el conjunt d'accions que es duen a terme amb l'objectiu de determinar, curar o pal·liar aquelles que no siguin la que requereix més atenció.

Procediment terapèutic Conjunt d'accions que es duen a terme amb l'objectiu de curar o pal·liar una malaltia.

Recurs sanitari Disponibilitats econòmiques per tal de garantir l'atenció, promoció i educació de la salut dels pacients en un centre sanitari.

Registre hospitalari Base de dades que emmagatzema les dades referents a les altes hospitalàries d'un o més hospitals.

Registre de tumors Base de dades que emmagatzema les dades referents a les localitzacions tumorals.

Registre hospitalari de tumors Base de dades que emmagatzema les dades referents als diagnòstics de tumors per a cada individu i les respectives altes hospitalàries relacionades.

Servei sanitari Conjunt d'activitats que conformen l'assistència sanitària en un centre sanitari.

Apèndix D

Acrònims

AC Departament d'Arquitectura de Computadors.

AH Base de Dades de les Altes Hospitalàries.

ASEDAT Aplicació de Sel·lecció i Extracció de Dades Tumorals.

APA Base de dades de l'Anatomia Patològica.

BSD *Berkeley Software Distribution.*

CMBD Conjunt Mínim Bàsic de Dades.

CSV *Comma Separated Value.*

DAMA-UPC *Data Managment Group* - Universitat Politècnica de Catalunya

ICD9 *International Classification of Diseases, Ninth Revision.*

ICO Institut Català d'Oncologia.

LOPD Llei Orgànica de Protecció de Dades.

PDO Pla Director d'Oncologia.

RHT Registre Hospitalari de Tumors de l'Hospital Universitari de Bellvitge i de l'Institut Català d'Oncologia.

UML *Unified Modeling Language.*

XML *eXtensible Markup Language.*

Bibliografia

[1] *Activitat assistencial de la xarxa sanitària de Catalunya: Informe del registre del CMBD.*
Generalitat de Catalunya, Departament de Sanitat i Seguretat Social
Servei català de la Salut, Barcelona, 2007.

[2] *Automated data collection in cancer registration.*
Black RJ, Siminato L, Storm HH, Démaret E
IARC, Lyon, 1998.

[3] *Automatization of a hospital-based tumor registry.*
Josepa Ribes, Jordi Gálvez, Àngels Melià, Ramon Clèries, Xavier Messeguer, Francesc Xavier Bosch
Gac Sanit., Barcelona, 2005.

[4] *Cancer registration. Principles and methods.*
Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG
IARC Scientific Publications, Lyon, 1991.

[5] *Composición de textos científicos con LaTeX.*
Gabriel Valiente Feruglio
Edicions UPC, Barcelona, 1997.

[6] *DEX: High-Performance Exploration on Large Graphs for Information Retrieval.*
Norbert Martínez Bazan, Victor Muntés Mulero, Sergi Gómez Villamor, Jordi Nin, Mario-A. Sánchez Martínez, Josep Lluís Larriba Pey
DAMA-UPC, Barcelona, 2007.

[7] *Enginyeria Del Software. Especificació de sistemes orientats a objectes amb la notació UML.*
Dolors Costal Costa, María Ribera Sancho Samsó, Ernest Teniente López
Edicions UPC, Barcelona, 2003.

[8] *International Statistical Classification of Diseases, Injuries, and Causes of Death. Ninth revision.*
Clinical Modification, US Department of Health and Human Services
Washington DC, 1979.

[9] *Manual de notificació per al Registre del CMBD: Hospitals Generals d'Aguts.*
Generalitat de Catalunya, Departament de Sanitat i Seguretat Social
Servei català de la Salut, Barcelona, 2003.

[10] *Pla Director d'Oncologia a Catalunya.*
Direcció General de Planificació i Avaluació
Departament de Salut - Generalitat de Catalunya, Barcelona, 2006.

[11] *Proyección de la incidencia y la mortalidad del càncer en Cataluña hasta el año 2015 mediante un modelo bayesiano.*
Josepa Ribes, Ramon Clèries, Maria Buxo, Alberto Ameijide, Joan Valls, Rosa Gispert
Med Clin, Barcelona, 2003.

[12] *SNOMED RT: A reference terminology for health care.*
Spackman KA, Campbell KE, Cote RA
College of American Pathologist, Northfield, IL, 2000.

[13] *BIBEX*
`http://www.dama.upc.edu/Bibex/`

[14] *CMBD*
`http://www10.gencat.net/catsalut/cat/prov_cmbd.htm`

[15] Commons Configuration - Java Configuration API
`http://commons.apache.org/configuration/`

[16] *Conjunt Mínim Bàsic de Dades (CMBD)*
`http://www10.gencat.net/catsalut/cat/prov_cmbd.htm`

[17] *DAMA-UPC*
`http://www.dama.upc.edu/`

[18] *Developer resources for Java Technology*
`http://java.sun.com/`

[19] *GraphML format*
`http://graphml.graphdrawing.org/`

[20] *International Classification of Diseases, 9th revision - ICD9*
`http://www.cdc.gov/nchs/about/otheract/icd9/abticd9.htm`

[21] JExcelApi - A Java API to read, write, and modify Excel spreadsheets
`http://jexcelapi.sourceforge.net/`

[22] LyX: The document processor based on TEX/LATEX
`http://www.lyx.org/`

[23] *Netbeans*
`http://www.netbeans.org/`

[24] *Prefuse - interactive information visualization toolkit*
`http://prefuse.org/`

[25] Registre Hospitalari de l'Hospital Universitari de Bellvitge i de l'Institut Català d'Oncologia (RHT)
`http://rht.iconcologia.catsalut.net/`

[26] yEd - Java Graph Editor
`http://www.yworks.com/`