



PROJECTE BIOINFORMÀTICA

Anàlisi en mortalitat del càncer

Relacions estadístiques

Aplicació que amb l'anàlisi del creuament de dades permet consultar i visualitzar dades sobre la mortalitat del càncer.

MATÈRIA: M12 Projecte

GRUP: 2T

DATA: 22/5/2022

Erik Joaquín Tingsvall

Jaime Sánchez Cárcel

Marcello Krell

Índex

Definició del projecte	3
Requeriments de Software	3
□ Front-end.....	3
□ Back-end	3
□ VCS.....	3
□ Solution Stack	3
□ Recomanacions	4
□ Cleanup Data	4
Anàlisi funcional.....	4
Disseny.....	4
□ Diagrama de Gantt.....	5
□ WireFrame	6
□ Diagrama Entitat-Relació	7
Desenvolupament del projecte	10
Proves i integració	10
Validació	10
Producció.....	10
Manteniment.....	10
Explicació i desenvolupament del projecte.....	11
Conclusions. Dificultats trobades. Possibles millores.	12

Definició del projecte

El funcionament de l'aplicació web és mostrar estadístiques amb relacions entre diferents tipus de dades:

- Morts per càncer amb els seus diferents tipus.
- Regions/països
- Edats
- Sexe

Creuar aquestes dades i treure conclusions basades en l'anàlisi de la informació recollida, processada/filtrada i mostrada amb diferents tipus de gràfiques.

Requeriments de Software

Aplicació web feta amb les següents tecnologies:

- **Front-end.**
 - HTML5
 - CSS3
 - JavaScript
- **Back-end.**
 - Model-Vista-Controlador en PHP
 - Base de dades amb MySQL
- **VCS** → Software de Control de Versions GitHub, amb dos repositoris:
 - [projectBigDataCancerDocumentation](#) → per la documentació.
 - [projectBigDataCancerSourceCode](#) → pel codi font de l'aplicació.
- **Solution Stack** → XAMPP, que inclou.
 - Servidor Apache.
 - Base de Dades MySQL amb gestor PhpMyAdmin.
 - PHP.

- **Recomanacions** per desenvolupar el projecte.
 - IDE de JetBrains PhpStorm, o editor de codi Visual Studio Code.
 - Per poder fer debug s'ha d'instal·lar → php_xdebug.dll
 - Ajuda per fer la instal·lació [Xdebug Installation](#)
- **Cleanup Data**, per fer neteja de les dades en cru (dataraw) fem ús d'Excel:
 1. Fórmula Excel per separar les causes de morts que contenen entre d'altres el càncer (C).
 - 1.1. =SI(IZQUIERDA(F2,1)="C","OK","KO")
 - 1.2. Ordenar descendentment
 - 1.3. Esborrar totes les files amb valor "KO".
 2. Per treure les morts únicament de càncer, son les que van fins el núm. 97.
 - 2.1. Dades → Text en Columnes → Delimitats → Separadors → C → Finalitzar
 - 2.2. Ordenar ascendentment.
 - 2.3. Esborrar els registres majors a 97.

Anàlisi funcional

- Objectius i abast del projecte: interpretar dades de mortalitat basades en el càncer lligats al sexe, l'edat, localització geogràfica.
- Resultats esperats: estudiar la variabilitat, analitzar i interpretar les dades en funció dels resultats de les gràfiques.
- Terminis establerts: 3 mesos.

Disseny

- Estimació d'hores: 330h

- **Diagrama de Gantt:**

BigDataCancer: Planning del treball del projecte

Fases del projecte	Setmana 1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
Fase inicial	Definició del projecte	Recerca de dades										
Fase 1			Diagrames: E/R, Classes, Casos d'Ús									
F2				WireFrame i Guia d'Estil	install.php (MySQL) HTML+CSS+JS							
F3						HTML+CSS+JS	HTML+CSS+JS	Cleaning DataRAW				
F4									HTML+CSS +JS+PHP	HTML+CSS +JS+PHP		
F5											HTML+CSS+JS +PHP+Chart.js	
Fase final												HTML+CSS+JS +PHP+Chart.js

Comentaris:

WireFrame:

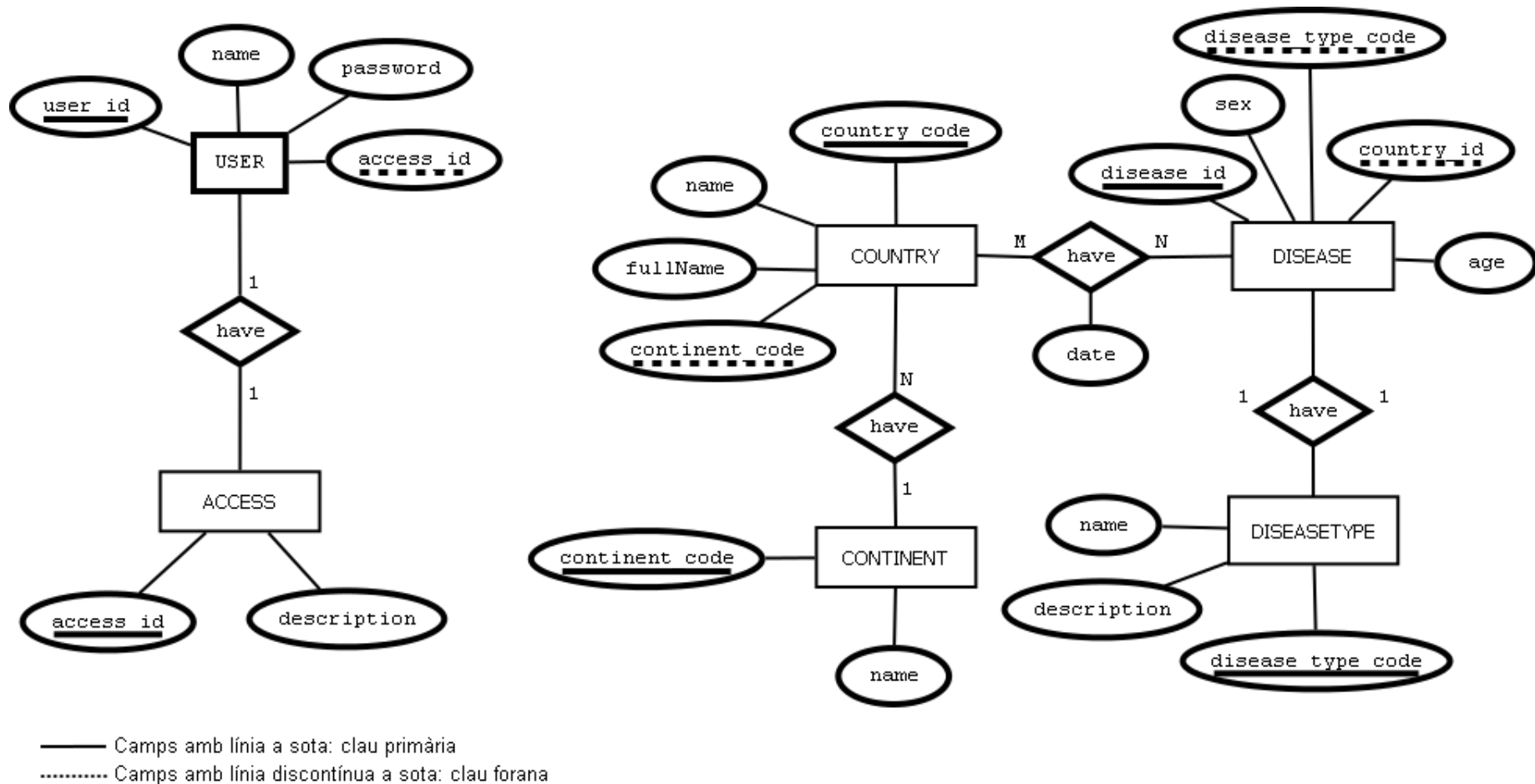
The wireframe illustrates the layout of the BigDataCancer website. Key components include:

- Header:** Features the BigDataCancer logo, a navigation menu (Causes-definició, Factors de risc, Registres, Recursos, Contacte), and a login section with fields for 'Usuari / Correu electrònic' and 'Clau', and a 'No s'ha identificat' message.
- Left Sidebar:** Contains filters for 'Alimentació deficient', 'Alcohol', 'Obesitat i falta d'exercici', 'Exposició a raigs UV', 'Drogues', and 'Contaminació'. It also includes a 'COMBO - Regions' dropdown and a 'COMBO - Països' dropdown.
- Main Content Area:** Displays a video player with a 'PETICIÓN' button, a 'RESPUESTA' button, and a 'CONTACTE' button. Below the video is a section titled 'Què és el càncer?' and 'Causes', followed by a 'FACTORS DE RISC DEL CàNCER' pie chart.
- Right Sidebar:** Includes a 'INICIAR SESIÓ' button, a 'BLOC CENTRAL' section with 'INFORMACIÓ DE LES PÀGINES' and 'MENÚ-SUBMENÚ', and a 'Normativa de Qualitat' section.
- Footer:** Contains contact information for 'OFICINAS' and 'PROYECTOS I+D', a 'HORARIO' section, and a 'Ver en Google Maps' button.

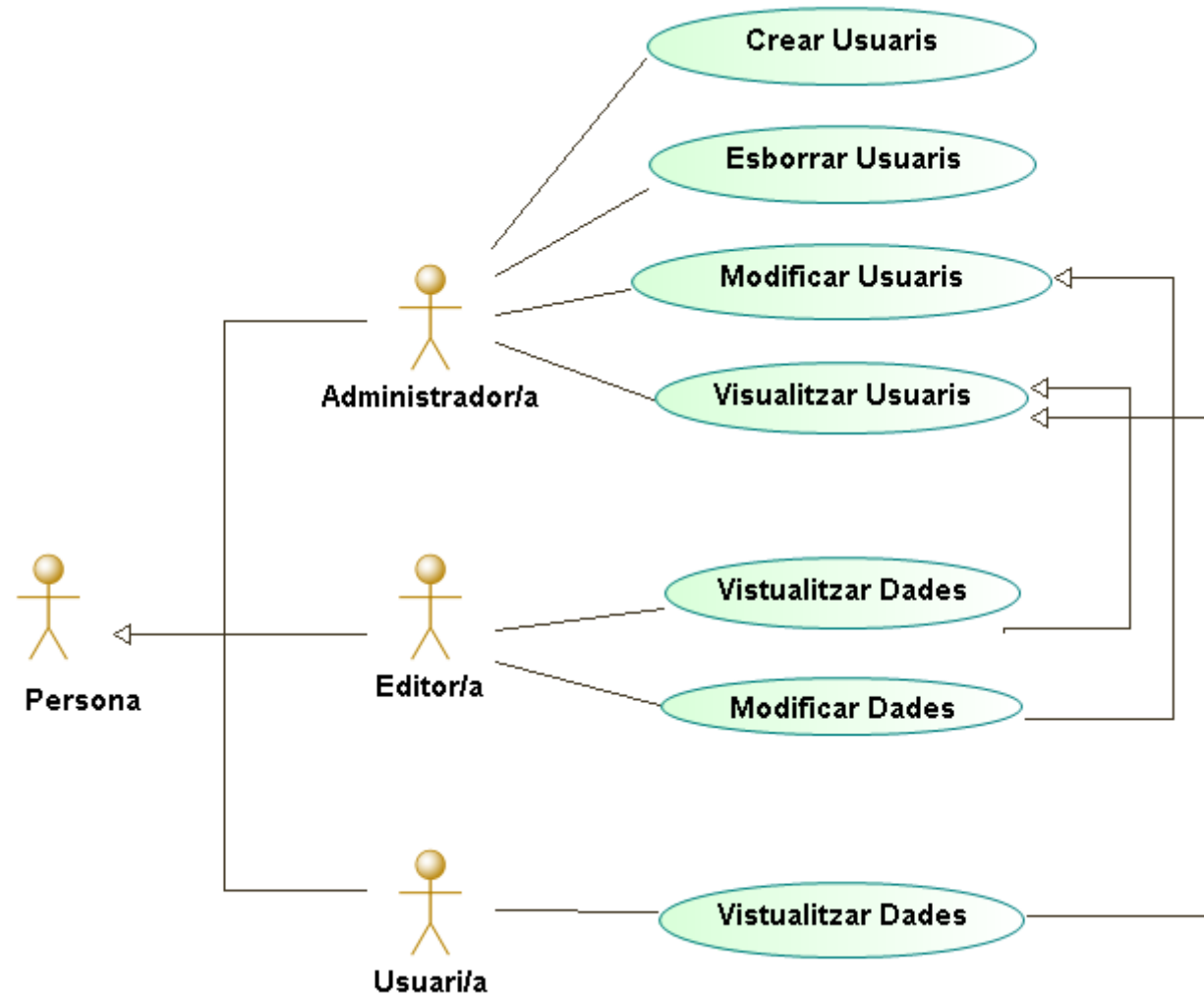
Annotations highlight specific interactions and elements:

- PETICIÓN:** Points to the video player and the 'CONTACTE' button.
- RESPUESTA:** Points to the 'RESPUESTA' button.
- PETICIÓN - DEJAR MENSAGE:** Points to the 'CONTACTE' button.
- COMBO - Regions:** Points to the 'COMBO - Regions' dropdown.
- COMBO - Països:** Points to the 'COMBO - Països' dropdown.
- CONTACTE:** Points to the 'CONTACTE' button.
- INICIAR SESIÓ:** Points to the 'INICIAR SESIÓ' button.
- BLOC CENTRAL:** Points to the 'BLOC CENTRAL' section.
- Normativa de Qualitat:** Points to the 'Normativa de Qualitat' section.

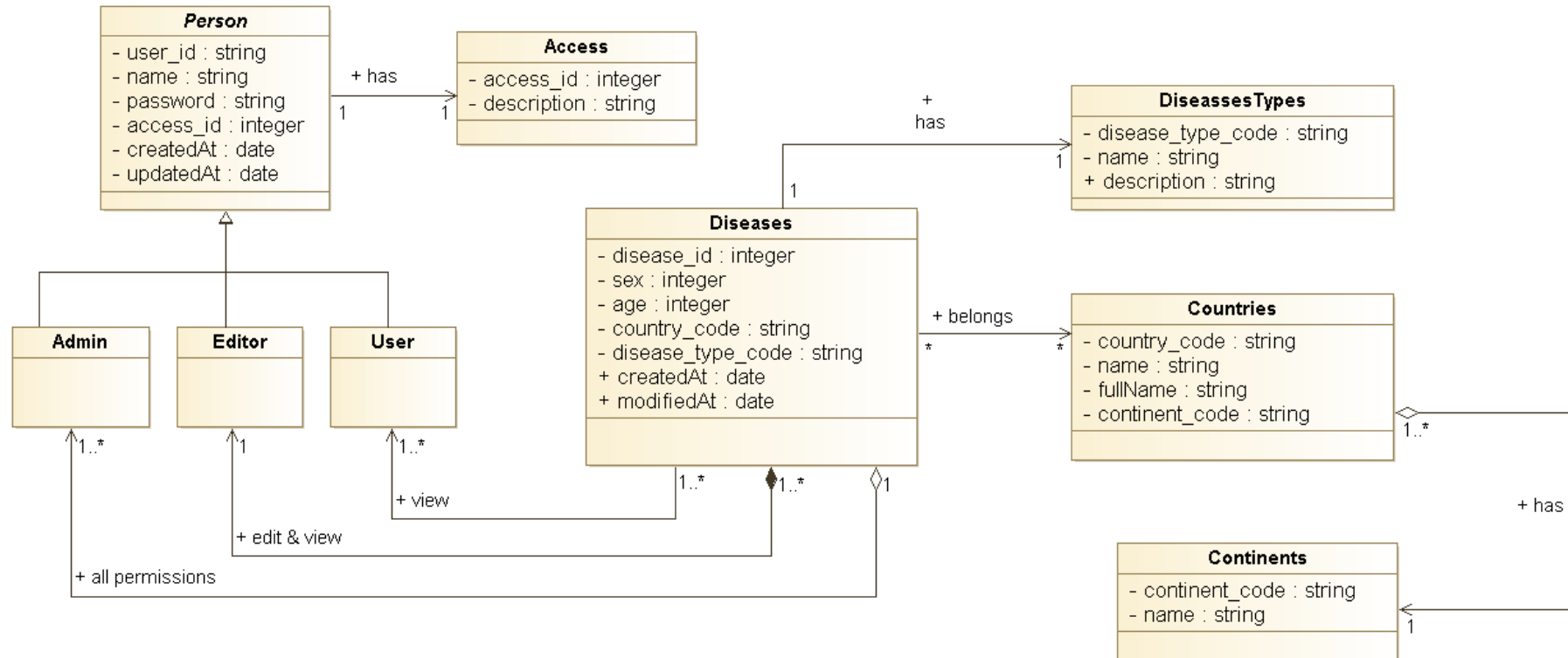
- Diagrama Entitat-Relació:



- Diagrama de casos d'ús.



- Diagrama de Classes



- Equip. Tres persones; Erik, Jaime i Marcello.

Desenvolupament del projecte

- Codificació / programació (codi font): traduir les funcionalitats i requeriments especificats en la fase d'anàlisi funcional.

Proves i integració

- Proves de validació: l'aplicació compleix les necessitats.
- Proves d'usuari: resultats esperats als usuaris.

Validació

Elaborar solucions a les necessitats expressades en el projecte. Validació de l'usuari conforme el lliurament del projecte, en relació amb les especificacions inicials.

Al final, s'elabora una documentació amb la finalitat de reunir la informació necessària per al seu ús i desenvolupaments posteriors.

Producció

Lliurament del producte final i posada en marxa. Informar als usuaris sobre les noves funcionalitats que ofereix la solució.

Manteniment

Possibles ajustos que incloguin tant accions correctives com evolutives:

Correcció de fallades de funcionament, errors tècnics, millores en l'experiència de l'usuari, (UX).

Requereix un seguiment, manteniment i millora de la solució, amb la finalitat de garantir la seva eficàcia i utilitat.

Explicació i desenvolupament del projecte.

El projecte és una pàgina web amb servidor XAMPP sobre estadístiques mundials del càncer, amb la que poder fer consultes, comparatives i gràfiques entre diferents països del món entre els anys 2003 i 2020.

Les dades han sigut extretes de la pàgina oficial de la Organitzación Mundial de la Salut (World Health Organization) i són de domini públic.

Les dades que podem trobar tenen 86 tipus de diagnòstics de càncer i 133 països fent distinció de sexe i edat en intervals de 5 fins a majors de 95 anys.

Primerament vam realitzar el funcionament de la pàgina amb un MVC (Model Vista Controlador) amb el que realitzar la gestió d'usuaris programat amb PHP.

Seguidament vam realitzar una primera versió de guia d'estil i WireFrame de la pàgina que vam anar modificant a mesura que necessitàvem adaptar funcionalitats com afegir les icones de xarxes socials, marges sensibles al grandària de la finestra del navegador per la seva correcta visualització i demés.

A continuació vam netejar la base dades. Volíem centrar-nos en una única causa de mort(càncer) però les dades amb les que treballàvem contenien totes les possibles causes de mort, així que vam filtrar per mitjà del codi ICD (International Classification of Diseases) les que tenien com a causa de mort la que pertanyia reduint el pes de l'arxiu de 80 MB a 8MB aproximadament.

Per últim vam integrar les dades amb SQL per a realitzar les consultes des de la web, fent proves i aplicant les restriccions adients per a evitar errors. Amb això ja només queda revisar la documentació produïda al llarg del projecte de cara a la seva exposició.

Conclusions. Dificultats trobades. Possibles millores.

Primerament ens hem donat compte de lo important que es posar les dades en comú per facilitar la seva explotació, per poder implementar la nostra base de dades em hagut de implementar el codi ISO 3166-1 per associar-lo a cada país ja que tenien una id pròpia.

Amb aquest projecte podem facilitar gràficament estadístiques sanitàries per a facilitar la recerca sobre morts per càncer amb els filtres que oferim apropant la tecnologia a la investigació.

La base de dades de la World Health Organization(WHO, OMS en català) tenia dades anteriors a l'any 2003 però no les hem utilitzar perquè contenien varies entrades per la mateixa malaltia dins del mateix any, vam provar de sumar el total de morts però no quadrava contrastant amb altres fonts ni teníem cap indicatiu per escollir cap de les entrades. Com passava de manera generalitzada vam descartar aquells anys.

No vam trobar dades de Rússia, de Xina les aportades corresponien al 10% de la població rural i no concretava les poblacions concretes, per tant les vam descartar també. Altres països descartats van ser: a la Palestina ocupada perquè oferia dades parcials, a Líban només aportava dades de població nacional i a Gabon utilitzaven un format de recompte de morts únic que dificultava la seva implementació, hauríem d'haver creat una excepció concreta que només s'aplicava a ells.

Volíem també aportar dades de la població total de cada país per a oferir una visió del ràtio de morts amb població, però l'arxiu de dades de la WHO de població tenia varies entrades per cada país per any, la suma de les quals tampoc corresponia amb la població total contrastada amb altres medis.

Com a possible millora podríem implementar una API de la base de dades que hem adaptat per a la creació del projecte i un formulari de pujada d'arxius per a engrandir la base de dades, però a falta de temps requerit no vam arribar a plantejar-ho.

