Graduado en Ingeniería de la salud

Proyecto Bases de Datos Biológicas

Realizado por
Alejandro Domínguez Recio

Tutorizado por
Ismael Navas Delgado

MÁLAGA, abril 2022

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA
Ingeniería de la Salud

Proyecto Bases de Datos Biológicas

GeneAssociations

Realizado por
Alejandro Domínguez Recio

Tutorizado por
Ismael Navas Delgado

UNIVERSIDAD DE MÁLAGA
MÁLAGA, JUNIO DE 2022

Resumen

Teniendo en cuenta la cantidad de datos asociados a genes hemos agrupado en la misma base de datos los más relevantes en un ámbito médico o de investigación. Para ello hemos creado una sistema gestor de base de datos en MySql server. A su vez se hará uso del paquete de consulta y extracción de datos Rentrez en R. La plataforma NCBI será la principal fuente de datos.

Índice

1

Introducción

1.1 Motivación

Actualmente se pueden consultar bases de datos relacionas a información genética tales como el propio NCBI, EMBL o DDBJ. Estas nos pueden proporcionar secuencias de un determinado gen, localización en el cromosoma o proteína codificante entre otros. Entre estos sistemas gestores de bases de datos podemos encontrar diversificaciones internas o referencias cruzadas entre ellas. Por todo esto cuando el personal investigador o médico necesita información de mas de una de estas, está obligado a realizar diferentes consultas independientes. La finalidad de GenAssociations no es aglutinar toda la información de todas estas en un sistema independitente si no crear un sistema gestor que permita reunir y proporcionar información sobre una serie de campos que están altamente relacionados facilitando y mejorando la expericiencia de los usuarios.

1.2 Objetivos

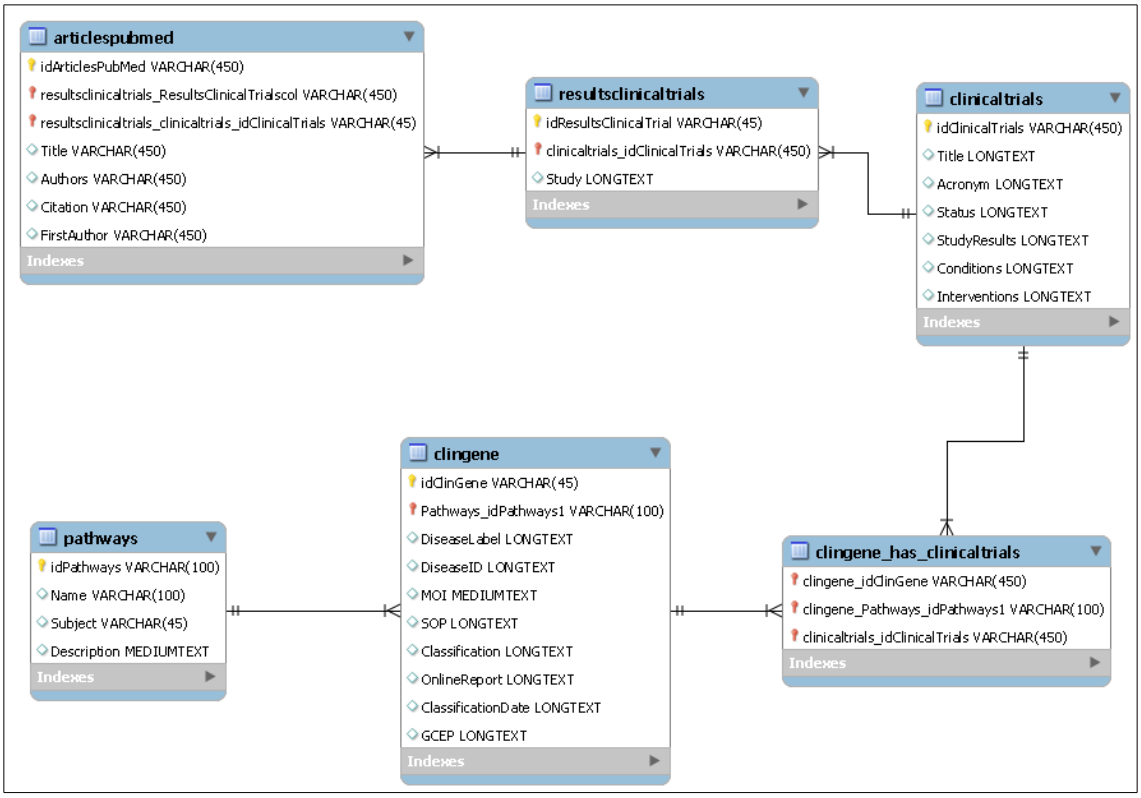
Como hemos comentado anteriormente nuestros principales objetivos es desarrollar un sistema gestor de base datos que facilite y mejore la experiencia de los usuarios respecto de las bases de datos existentes. Para ello consideramos fundamental que las consultas proporcionas cubran las posibles necesidades de los usuarios. La creación de las consultas estará guiada por la lógica de relación que hay detrás de esta por lo que deberá ser cuidada especialmente. Un paso anterior a esto será la introducción de los datos la cual será basada en lógica biológica. Por lo que recapitulando nuestros principales objetivos serán un conjunto de consultas que cubran las diferentes necesidades de los usuarios, una lógica cuidada que guíe la relación de los diferentes elementos del sistema gestor de datos y evite conflictos entre estos.

2

Descripción Esquema Base de Datos

2.1 Esquema base de datos

En el siguiente esquema se puede ver el modelado de las tablas, atributos y relaciones que forman nuestra base de datos. La lógica del modelo ha sido guiada por la representación real de los distintos componentes y sus respectivas dependencias.



2.1.2 Descripción de tablas

- **Pathways.**

  Representa las distintas rutas metabólicas dentro de los humanos.

- **ClinGene**

  Representa los distintos genes involucrados en rutas metabólicas.

- **ClinicalTrials**

  Representa estudios clínicos sobre genes.

- **ResultsClinicalTrials**

  Representa los resultados de los estudios clínicos.

- **ArticlesPubmed**

  Representa artículos científicos sobre los resultados de estudios clínicos.

2.1.3 Descripción de relaciones

- **Pathways - ClinGene**

  Se tiene en cuenta que un gen solo puede estar asociado a una ruta metabólica y que una ruta metabólica puede estar formada por muchos genes. Un gene tiene que tener una ruta metabólica para su presencia.

- **ClinGene – ClinicalTrials**

  Cada estudio clínica deberá de tener asociado un gen y por lo tanto su ruta metabólica. La existencia de un estudio clínico es posterior a la presencia del gene y su respectiva ruta metabólica.

- **ClinicalTrials – ResultsClinicalTrials**

  Cada estudio clínico deberá de tener asociado un resultado.

- **ResultsClinicalTrials - ArticlesPubmed**

  Cada artículo científico deberá de estar asociado un único resultado de estudio clínico. Por el contrario un resultado de estudio clínico podrá tener asociado varios artículos científicos

2.2 Procedencia de los datos

Siendo la intención del proyecto facilitar una visión amplia sobre la información biológica sanitaria asociada a una patología, se han agrupado datos de diversas fuentes. Las fuentes a partir de las cuales hemos creado nuestra base de datos son NCBI (The National Center for Biotechnology Information) y NIH ClinicalTrials. Los disntintos conjuntos de datos han sido modificados mínimamente para ser adaptados a las características propias del esquema de base de datos.

3

Diseño de consultas

3.1 Objetivo de las consultas

Las consultas serán diseñadas con fin de modelar situaciones en las que se requiera obtener información de la base datos. Se modelan situaciones lo más cercanas a las necesidades de los usuarios de la base de datos. Exclusivamente se crearán consultas de selección. En primera instancia no se tendrá en cuenta técnicas de optimización de consultas tales como la creación de índices o cambio de motor de búsqueda.

3.2 Consultas

El conjunto de consultas creadas irán diferenciadas en base a los comandos SQL utilizados.

3.2.1 Consultas Where

Las consultas where tienen la función del filtrado de resultados, dada una condición.

3.2.1.1 Consulta Where 1

Obtener las descripciones de las rutas metabólicas cuyo subject = 'Metabolic

```
select Description from pathways where Subject = 'Metabolic' limit 150;
```

Description
Alanine (L-Alanine) is an ?-amino acid that is used for protein bio...
Aspartate is synthesized by transamination of oxaloacetate by ...
Glutamate is one of the non-essential amino acids that is produc...
Glutathione (GSH) is an low-molecular-weight thiol and antioxida...
The citric acid cycle, which is also known as the tricarboxylic acid...
Linoleic acid (LNA) is a polyunsaturated fatty acid (PUFA) precur...
Selenoamino acids include selenocysteine, selenohomocysteine ...
Amino sugars are sugar molecules containing an amine group. T...
Ammonia can be rerouted from the urine and recycled into the b...
The arginine and proline metabolism pathway illustrates the bios...
Beta-alanine, 3-aminopropanoic acid, is a non-essential amino a...
Betaine (or trimethylglycine) is similar to choline (trimethylamino...
Biotin is a vitamin that is an essential nutrient for humans. Biotin...

3.2.1.2 Consulta Where 2

Obtner el subject de todas las rutas metabólicas en cuya descripción contenga 'citric'

```
select Subject from pathways_copy1 where Description like '%Citric%';
```

Subject
Metabolic
Metabolic
Metabolic
Metabolic
Metabolic
Metabolic
Metabolic
Metabolic
Disease
Disease
Disease
Disease

3.2.1.3 Consulta Where 3

Obtener el número de estudios clínicos donde las intervenciones son 'Drug Perifosine|Capecitabine'

```
select count(Conditions) from clinicaltrials where Interventions = 'Drug: Perifosine|Drug: Capecitabine';
```

Resultados

count(Conditions)
1

3.2.2 Consultas Where + Inner Join

Las consultas Join tienen la función de combinar dos o más tablas, basándose en la relación de sus columnas.

3.2.2.1 Consulta Where + Inner Join 1

Obtener las citaciones de un artículos Pubmed los cuales estén asociados a un estudio clínico a 'Dexamethasone'

```
select Citation
from articlespubmed
inner join resultsclinicaltrials
on resultsclinicaltrials.idResultsClinicalTrial = articlespubmed.resultsclinicaltrials_idResultsClinicalTrials
where Study
like '%Dexamethasone%';
```

Resultados

Citation
Cancer Res. 2009 Jul 1;69(13):5269-84. doi: 1...

3.2.2.2 Consulta Where + Inner Join 2

Obtener el tipo de estudio asociado a unos resultados de un estudio clínico en los cuales se haga referencia a 'Update' y 'Autoimmune'.

```
select Study
from resultsclinicaltrials
inner join articlespubmed
on  resultsclinicaltrials.idResultsClinicalTrial = articlespubmed.resultsclinicaltrials_idResultsClinicalTrials
where Title
like '%Update%'
and Title
like '%Autoimmune%';
```

Study
study interventions are Thalidomide . kidney cancer diagnosis and no diabetes mellitus

3.2.2.3 Consulta Where + Inner Join 3

Obtener el nombre, descripción, nombre de enfermedad y GCEP asociados a patologías metabólicas con la etiqueta 'Perixomal Disorders.

```
select Name, Description, DiseaseLabel, GCEP
from pathways
inner join clingene
on  pathways.idPathways = clingene.Pathways_idPathways1
where GCEP = 'Peroxisomal Disorders';
```

Resultados

Name	Description	DiseaseLabel	GCEP
Alpha Linolenic Acid and Linoleic Acid Metabolism	Linoleic acid (LNA) is a polyunsaturated fatty acid (PUFA) precur…	adrenoleukodystrophy	Peroxisomal Disorders
Selenoamino Acid Metabolism	Selenoamino acids include selenocysteine, selenohomocysteine …	congenital bile acid synthesis defect 5	Peroxisomal Disorders
Degradation of Superoxides	Reactive oxygen species (ROS) are formed by the normal meta…	acyl-CoA binding domain containing protein 5 de…	Peroxisomal Disorders
Ethanol Degradation	Ethanol metabolism in humans occurs mainly in the liver, though …	peroxisomal acyl-CoA oxidase deficiency	Peroxisomal Disorders
Lactose Synthesis	Lactose synthesis occurs only in the mammary glands, producin…	alkylglycerone-phosphate synthase deficiency	Peroxisomal Disorders
Valine, Leucine, and Isoleucine Degradation	Valine, isoleuciine, and leucine are essential amino acids and are…	alanine glyoxylate aminotransferase deficiency	Peroxisomal Disorders
Lesch-Nyhan Syndrome (LNS)	Lesch-Nyhan Syndrome (LNS; Hypoxanthin guanine phosphorib…	alpha-methylacyl-CoA racemase deficiency	Peroxisomal Disorders
Sphingolipid Metabolism	The sphingolipid metabolism pathway depicted here describes th…	bile acid CoA:amino acid N-acyltransferase defic…	Peroxisomal Disorders
Lisinopril Action Pathway	Benazepril, brand name Lotensin, belongs to the class of drugs …	acatalasia	Peroxisomal Disorders

3.2.3 Consultas Where + Subqueries

Las consultas where + subqueries marcarán registros de una tabla existentes en una subquery.

3.2.3.1 Consulta Where + Subquery 1

Obtener todas las intervenciones aplicadas en un estudio clínico en las cuales el resultado de estudio contiene la etiqueta 'cancer'

```
select Interventions
from clinicaltrials
where exists
(select study
from resultsclinicaltrials
where resultsclinicaltrials.clinicaltrials_idClinicalTrials = clinicaltrials.idClinicalTrials
and Study like '%cancer%');
```

Interventions
Procedure: sentinel lymph node mapping
Procedure: robotic assisted surgery|Procedure:...
Procedure: Radical resection of colon cancer
Procedure: laparoscopic conventional colectomy...
Procedure: hvNOTES radical colectomy
Procedure: Experimental group|Procedure: Con...
Procedure: D2 radical operation|Procedure: Co...
Procedure: Conventional Surgery|Procedure: C...
Procedure: Conventional Right hemicolectomy (...
Procedure: Chiropractic high velocity low amplit...
Procedure: blue and isotopic detection of sentin...
Other: Primary Care Provider/Staff Participants ...
Other: Moderate intensity physical effort|Other...
Other: Enhanced recovery program

3.2.3.2 Consulta Where + Subquery 2

Obtener tdos los títulos de los artículos pubmed en los cuales la condición tratada en el estudio clínico asociado es 'Colon Cancer'

```
select Title
from articlespubmed
where resultsclinicaltrials_clinicaltrials_idClinicalTrials in
( select idClinicalTrials
from clinicaltrials
where Conditions = 'Colon Cancer' );
```

Title
Cancer prevention: from 1727 to milestones of the past 100 years
Global Cancer Incidence and Mortality Rates and Trends--An Update
Measuring cancer evolution from the genome
Clinical, Prognostic and Therapeutic Significance of Heat Shock Proteins in Cancer
Cancer-associated fibroblasts in tumor microenvironment - Accomplices in tumor ...
Tumor microenvironment: recent advances in various cancer treatments
What Is Cancer?
Addressing cancer's grand challenges

3.2.3.3 Consulta Where + Subquery 3

Obtener todas las etiquetas de enfermedad cuyo subject es 'Disease' y la descripción de la ruta metabólica contiene 'Bile Acid'

```
select DiseaseLabel
from clingene
where Pathways_idPathways1 in
( select idPathways
from pathways
where Subject = 'Disease' and Description like '%Bile Acid%');
```

Resultados

DiseaseLabel
mosaic variegated aneuploidy syndrome 1
C1Q deficiency
frontotemporal dementia and/or amyotrophic la...
autosomal dominant distal hereditary motor neu...

4

Optimización Base de Datos

4.1 Objetivo de la optimización

Con la optimización de consultas pretendemos ver el efecto que tiene la elección del motor de búsqueda asociado a las diferentes tablas como las incorporación de índices. Se mostrará de forma a modo de tabla las respectivas comparaciones.

4.2 Consultas optimizadas

4.2.1 Consultas Where

Tiempo de ejecución

Consulta	InnoDB	MyISAM	Índice + InnoDB	Índice + MyISAM
1	0,0122	0,03103	0,00875	0,008994
2	0,09424	0,15454	0,09591	0,152838
3	0,00122	0,00707	0,00182	0,008758

Filas inspeccionadas

Consulta	InnoDB	MyISAM	Índice + InnoDB	Índice + MyISAM
1	150/773	150/77	150/150	150/150
2	46/45557	46/4667	46/4557	46/4557
3	1/100	1/101	1/1	1/1

4.2.2 Consultas Where + Inner Join

Tiempo de ejecución

Consulta	InnoDB	MyISAM	Índice + InnoDB	Índice + MyISAM
1	0,005343	0,00494	0,001081	0,01227
2	0,000735	0,00985	0,000626	0,01239
3	0,011787	0,12406	0,002068	0,10448

Filas inspeccionadas

Consulta	InnoDB	MyISAM	Índice + InnoDB	Índice + MyISAM
1	1/24	1/110	1/24	1/110
2	1/11	1/110	1/11	1/110
3	34/1712	34/6235	34/64	34/4591

4.2.3 Consultas Where + Subquery

Tiempo de ejecución

Consulta	InnoDB	MyISAM	Índice + InnoDB	Índice + MyISAM
1	0,01364	0,01255	0,01190	0,01421
2	0,00849	0,00883	0,00767	0,00751
3	0,05866	0,10993	0,01865	0,01717

Filas inspeccionadas

Consulta	InnoDB	MyISAM	Índice + InnoDB	Índice + MyISAM
1	38/176	38/239	38/176	38/239
2	8/20	8/119	8/20	8/97
3	4/4561	4/6239	4/4561	4/1912

La selección de los índices a ido asociada a la columna que debe cumplir la cláusula where. Los motores de búsqueda elegidos tienen la finalidad de ver el efecto del control referencial o de transacciones. En este caso InnoDB es un motor de búsqueda con control referencial y de transacciones por el contrario de MyISam. Este último tiene la ventaja de proporcionar mayor velociada a la hora de recuperar los datos. Esto lo hemos ido viendo en las distintas consultas.

çAunque en algunos casos la creación de índices asociados a las claves primarias o secundarias en InnoDB proporciona un rendimiento óptimo en comparación con MyISam. En el contexto de nuestra aplicación predominarán los selects por lo que MyISam podría ser mas aconsejable.

5

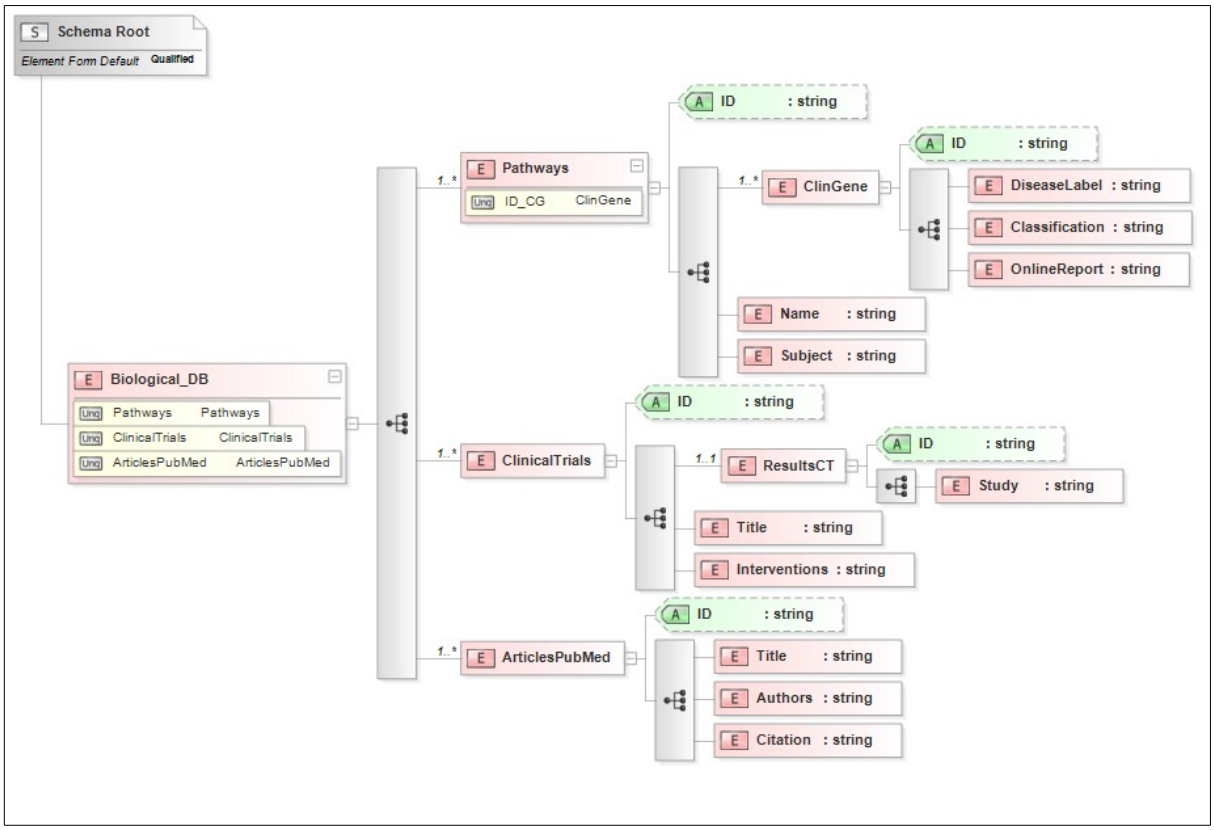
Modelo XML

5.1 Modelo XML

XML es un Lenguaje de Marcas Extensibles o meta-lenguaje. Esta formado por etiquetas que contienen la información que queremos almacenar y procesar. A la hora de elegir entre un modelo de datos relacional o XML se deben de tener en cuenta factores tales como la flexibilidad del modelo a representar (capacidad de modificación del modelo), rendimiento de la recuperación de datos (modelos relacionales suelen tener mejor rendimiento), procesado de los datos (rendimiento mayor modelos relacionales), presencia de atributos (requisitos de existencia de atributos menor en modelos XML), integridad referencia (XML no puede definir columnas como restricciones referenciales) o actualización del modelo (actualizaciones grandes aconsejable modelos relacionales).

Los documentos XML tienen asociados un esquema XSD, el cual comprueba la validez de su estructura, tipo de datos, atributos o orden. Para la desarrollo del documento XSD se ha utilizado el editor gráfico de documentos XSD y XML Liquid Studio.

5.1.1 Esquema XSD



El esquema XSD representa la estructura que soportará la información de nuestro conjunto de datos como también proporcionará la guía de como esta tiene que ser consultada.

Se ha seguido la recomendación de la literatura en evitar que la información sea guardada en atributos debido al aumento de la complejidad de lectura y mantenimiento. Los atributos se ha utilizado para guardar información no relevante sobre los datos como es son identificadores. Por consiguiente son los elementos los que contienen toda la información del conjunto de datos.

5.2 Generación de datos

La inserción de datos se ha llevado a cabo a partir del paquete elementTree de Python. Se han tenido en cuenta los diferentes tags y atributos de los distintos elementos del arbol y a partir de las característica de cada uno de ellos se ha ido incorporando los datos.

Fragmentos del código utilizado para insertar datos :

```
## Tabla Pathways
i = 0;
for elem in root.findall('Pathways'):
        elem.attrib['ID'] = df_Pathways['SMPDB ID'][i]
        elem.find('Name').text = df_Pathways['Name'][i]
        elem.find('Subject').text = df_Pathways['Subject'][i]
        i = i+1
        for iter in elem:
            print(iter.text)

## Tabla ClinGene
for elem in root.findall('Pathways'):
    for subelem in elem.findall('ClinGene'):
        subelem.attrib['ID'] = df_Clingen['ID'][i]
        subelem.find('DiseaseLabel').text = df_Clingen['DISEASE LABEL'][i]
        subelem.find('Classification').text = df_Clingen['CLASSIFICATION'][i]
        subelem.find('OnlineReport').text = df_Clingen['MOI'][i]
        i = i+1
        for iter in subelem:
            print(iter.text)

## Tabla ClinicalTrials
i = 0;
for elem in root.findall('ClinicalTrials'):
        elem.attrib['ID'] = str(df_clinicaltrials['NCT Number'][i])
        elem.find('Title').text = df_clinicaltrials['Title'][i]
        elem.find('Interventions').text = df_clinicaltrials['Study Type'][i]
        i = i+1
        for iter in elem:
            print(iter.text)

## Tabla ArticlesPubMed
i = 0;
for elem in root.findall('ArticlesPubMed'):
        elem.attrib['ID'] = str(df_cancer['PMID'][i])
        elem.find('Title').text = df_cancer['Title'][i]
        elem.find('Authors').text = df_cancer['Authors'][i]
        elem.find('Citation').text = df_cancer['Citation'][i]
        i = i+1
        for iter in elem:
            print(iter.text)

## Tabla ResultsClinicalTrials
i = 0;
for elem in root.findall('ClinicalTrials'):
    for subelem in elem.findall('ResultsCT'):
        subelem.attrib['ID'] = df_ResultsClinicalTrials['ID_Results'][i]
        subelem.find('Study').text = df_ResultsClinicalTrials['Study'][i]
        i = i+1
        for iter in subelem:
            print(iter.text)
```

5.2.1 Consultas Xquery

5.2.1.1 Consulta XQuery 1

Obtener todas los elementos ClinGene clasificados como 'Moderate'

```
for $elem in
doc("/db/XML_ProyectoFinal.xml")/Biological_DB/Pathways/ClinGene
where $elem/Classification = 'Moderate'
return $elem
```

Resultados

<ClinGene ID="ABCC4">

<DiseaseLabel>qualitative platelet defect</DiseaseLabel>
<Classification>Moderate</Classification>

<OnlineReport>AR</OnlineReport>

</ClinGene>

5.2.1.2 Consulta XQuery 2

Obtener todos los títiulos de ensayos clínicos que coincidan con un ID.

```
for $elem in
doc("/db/XML_ProyectoFinal.xml")/Biological_DB/ClinicalTrials
where $elem/@ID="NCT04031963"
return $elem/Title
```

Resultados

<Title>Novel Biophotonics Methodology for Colon Cancer Screening</Title>

5.2.1.3 Consulta XQuery 3

Obtener todos los títulos de ensayos clínicos ordenados descendentemente.

```
distinct-values(
for $elem in
doc("/db/XML_ProyectoFinal.xml")/Biological_DB/ClinicalTrials
order by $elem/Title
return data($elem)
)
```

Resultados

```
1
" study interventions are recombinant CD40-ligand . melanoma skin diagnosis and no active
cns metastases by ct scan or mri Colon Cancer Surgery in the Aged; Postoperative Outcome,
Functional Recovery and Survival. Observational "
2
" study interventions are Liposomal doxorubicin . colorectal cancer diagnosis and
cardiovascular Novel Biophotonics Methodology for Colon Cancer Screening Observational "
3
" study interventions are BI 836909 . multiple myeloma diagnosis and indwelling central
venous cateder or willingness to undergo intra venous central line placement The Efficacy
Of Complete Mesocolic Excision With Central Vessel Ligation Technique On Lymph Nodes And
Safety Margins Compared With Conventional Surgery For Colon Cancer Treatment Interventional
"
```

5.2.1.4 Consulta XQuery 4

Contar el numero de elementos ClinGene con OnlineReporte ='AR'

```
count(
for $elem in
doc("/db/XML_ProyectoFinal.xml")/Biological_DB/Pathways/ClinGene
where $elem/OnlineReport = 'AR'
return $elem
)
```

Resultados

```
1
10
```

5.2.1.5 Consulta XQuery 5

Obtener en formato HTML todos los autores cuyos articulos contengan la palabra 'Cancer'

```
<output>
    {
    for $elem in
doc("/db/XML_ProyectoFinal.xml")/Biological_DB/ArticlesPubMed
```

```
        where contains($elem//Title, "Cancer")
        return <Authors>{data($elem//Authors)}</Authors>
    }
    </output>
```

Resultados

<output>

<Authors>Hausman DM.</Authors>

<Authors>Torre LA, Siegel RL, Ward EM, Jemal A.</Authors>

<Authors>Lippman SM, Hawk ET.</Authors>

</output>

5.2.1.5 Consulta XQuery 6

Obtener en formato HTML todos los IDs de los elementos ClinGene cuyo DiseaseLabel contenga 'acyl-CoA', OnlineReport = 'AR y Classification = 'Definitive'

```
<output>
    {
    for $elem in
doc("/db/XML_ProyectoFinal.xml")/Biological_DB/Pathways/ClinGene
        where contains($elem//DiseaseLabel, "CoA")
        and $elem//OnlineReport = 'AR'
        and $elem//Classification = 'Definitive'
        return <ID>{data($elem//@ID)}</ID>
    }
    </output>
```

Resultados

<output>

<ID>ACAD8</ID>

<ID>ACAD9</ID>

<ID>ACADM</ID>

<ID>ACADS</ID>

</output>

6

Descripción MongoDB

6.1 Esquema base de datos

MongoDB es un sistema de base de datos NoSQL orientado a documentos. Los datos no son almacenados en tablas o registras, estos son almacenados en documentos y están representados a modo de colecciones. El formato de almacenamiento esta soportado por documentos BSON, la representación binaria de JSON.

Una de las diferencias más significativas respecto de las bases de datos relacionales o documentos XML es que en este sistema de almacenamiento no es necesario seguir un esquema.

A continuación procederemos a mostrar el conjunto de colecciones que hemos creado con el fin de dar soporta al conjunto de consultas realizadas anteriormente.

6.2 Generación de los datos

Para la generación de los datos aportados a la base de datos NoSQL MongoDB hemos utilizados las consultas de MySql. Concretamente hemos utilizados consultas con cláusulas INNER JOIN con el objetivo de unir varias tablas tablas y poder relacionarlas.

Como hemos mencionado anteriormente los datos son almacenados en documentos BSON aunque previamente a transformarlos en dicho formato es posible importarlos en CSV o JSON.

En nuestro caso por la posesión inicial de la mayoría del conjunto de datos en formato CSV se han importado en este formato.

7

Diseño de consultas MongoDB

7.1 Consultas

Debido al formato de los documentos de almacenamiento las consultas se hacen pasando objetos JSON como parámetros en MongoDB.

7.2 Consultas Where

7.2.1 Consulta Where 1

SQL

select Description from pathways where Subject = 'Metabolic' limit 150;

JSON

{Subject:"Metabolic"},{Description:1}

Resultados

	_id ObjectId	Description String
1	ObjectId('62bb45bf1ec5ce9c85e6…	"The metabolism of choline con…
2	ObjectId('62bb45bf1ec5ce9c85e6…	"Inositol phosphates are a gro…
3	ObjectId('62bb45bf1ec5ce9c85e6…	"Inositol (also known as myo-i…
4	ObjectId('62bb45bf1ec5ce9c85e6…	"The lipoic acid metabolism in…
5	ObjectId('62bb45bf1ec5ce9c85e6…	"The N-glycan biosynthesis is …
6	ObjectId('62bb45bf1ec5ce9c85e6…	"The metabolism of nitrogen in…
7	ObjectId('62bb45bf1ec5ce9c85e6…	"The riboneogenesis pathway is…
8	ObjectId('62bb45bf1ec5ce9c85e6…	"L-malic acid is metabolized t…
9	ObjectId('62bb45bf1ec5ce9c85e6…	"The TCA cycle (tricarboxylic …
10	ObjectId('62bb45bf1ec5ce9c85e6…	"The alanine biosynthesis star…
11	ObjectId('62bb45bf1ec5ce9c85e6…	"The metabolism of arginine be…
12	ObjectId('62bb45bf1ec5ce9c85e6…	"The metabolism of proline beg…

7.2.2 Consulta Where 2

SQL

select Subject from pathways where Description like '%Citric%';

JSON

{ Description : /Citric/},{Subject: 1}

Resultados

	_id ObjectId	Subject String
1	ObjectId('62bb45bf1ec5ce9c85e6…	"Metabolic"
2	ObjectId('62bb45bf1ec5ce9c85e6…	"Metabolic"

7.2.2 Consulta Where 3

SQL

select count(Conditions) from clinicaltrials where Interventions = 'Drug: Perifosine|Drug: Capecitabine';

JSON

{ Interventions : 'Drug: Perifosine|Drug: Capecitabine'}, {Conditions: 1}

Resultados

```
_id: ObjectId('62bc73effb5bcaba8d01cf9a')
Conditions: "Colon Cancer"
```

7.3 Consultas Where + Inner Join

7.3.1 Consultas Where + Inner Join 1

SQL

```
select Citation
from articlespubmed
inner join resultsclinicaltrials
on resultsclinicaltrials.idResultsClinicalTrial =
articlespubmed.resultsclinicaltrials_idResultsClinicalTrials
where Study
like '%melanoma%';
```

JSON

```
{Study:/melanoma/}, {Subject: 1}
```

*Tenemos en cuenta que la colleción de ha creado a partir de un join en Mysql.

Resultados

	_id ObjectId	Citation String
1	ObjectId('62bc74b0fb5bcaba8d01…	"Perspect Biol Med. 2019;62(4)…

Citation
Perspect Biol Med. 2019;62(4):778-784. doi: 10..

Se verifica que coinciden los resultados en MySql.

7.3.2 Consultas Where + Inner Join 2

SQL

```
select Study
from resultsclinicaltrials
inner join articlespubmed
on  resultsclinicaltrials.idResultsClinicalTrial =
articlespubmed.resultsclinicaltrials_idResultsClinicalTrials
where Title
like '%Update%'
and Title
like '%Autoimmune%';
```

JSON

```
{Title:/Update/, Title:/Autoimmune/},{Study:1}
```

*Tenemos en cuenta que la colleción de ha creado a partir de un join en Mysql.

Resultados

	_id ObjectId	Study String
1	ObjectId('62bc74b0fb5bcaba8d01…	"study interventions are Thali…

7.3.3 Consultas Where + Inner Join 3

SQL

```
select Name, Description, DiseaseLabel, GCEP
from pathways
inner join clingene
on  pathways.idPathways = clingene.Pathways_idPathways1
where GCEP = 'Peroxisomal Disorders';
```

JSON

```
{GCEP:'Peroxisomal Disorders'},{Name:1, Description:1, DiseaseLabel:1, GCEP:1}
```

*Tenemos en cuenta que la colleción de ha creado a partir de un join en Mysql.

Resultados

	_id ObjectId	Name String	Description String	DiseaseLabel String	GCEP String
1	ObjectId('62bc74d0fb5bcaba8d01…	"Alpha Linolenic Acid and Lino…	"Linoleic acid (LNA) is a poly…	"adrenoleukodystrophy"	"Peroxisomal Disorders"
2	ObjectId('62bc74d0fb5bcaba8d01…	"Selenoamino Acid Metabolism"	"Selenoamino acids include sel…	"congenital bile acid synthesi…	"Peroxisomal Disorders"
3	ObjectId('62bc74d0fb5bcaba8d01…	"Degradation of Superoxides"	"Reactive oxygen species (ROS)…	"acyl-CoA binding domain conta…	"Peroxisomal Disorders"
4	ObjectId('62bc74d0fb5bcaba8d01…	"Ethanol Degradation"	"Ethanol metabolism in humans …	"peroxisomal acyl-CoA oxidase …	"Peroxisomal Disorders"

7.4 Consultas Where + Subqueries

7.4.1 Consultas Where + Subqueries 1

SQL

```
select Interventions
from clinicaltrials
where exists
(select study
from resultsclinicaltrials
where resultsclinicaltrials.clinicaltrials_idClinicalTrials = clinicaltrials.idClinicalTrials
and Study like '%cancer%');
```

JSON

{Study:/cancer/}, {Interventions:1}

*Tenemos en cuenta que la colleción de ha creado a partir de un join en Mysql.

Resultados

	_id ObjectId	Interventions String
1	ObjectId('62bc74fbfb5bcaba8d01…	"Behavioral: Decision Aid for …
2	ObjectId('62bc74fbfb5bcaba8d01…	No field
3	ObjectId('62bc74fbfb5bcaba8d01…	"Colon Cancer"
4	ObjectId('62bc74fbfb5bcaba8d01…	"Drug: AquaminÂ®|Drug: Calcium…
5	ObjectId('62bc74fbfb5bcaba8d01…	"Procedure: Radical resection …
6	ObjectId('62bc74fbfb5bcaba8d01…	"Other: Moderate intensity phy…
7	ObjectId('62bc74fbfb5bcaba8d01…	No field
8	ObjectId('62bc74fbfb5bcaba8d01…	"Drug: Capecitabine|Drug: Oxal…
9	ObjectId('62bc74fbfb5bcaba8d01…	"Procedure: D2 radical operati…
10	ObjectId('62bc74fbfb5bcaba8d01…	"Procedure: laparoscopic conve…
11	ObjectId('62bc74fbfb5bcaba8d01…	"Procedure: Experimental group…
12	ObjectId('62bc74fbfb5bcaba8d01…	"Drug: Vinorelbine Tartrate"
13	ObjectId('62bc74fbfb5bcaba8d01…	No field
14	ObjectId('62bc74fbfb5bcaba8d01…	"Drug: Peri-operative chemothe…

7.4.2 Consultas Where + Subqueries 2

SQL

```
select Title
from articlespubmed
where resultsclinicaltrials_clinicaltrials_idClinicalTrials in
( select idClinicalTrials
from clinicaltrials
where Conditions = 'Colon Cancer' );
```

JSON

```
{Conditions:'Colon Cancer'}, {Title:1}
```

*Tenemos en cuenta que la colleción de ha creado a partir de un join en Mysql.

Resultados

	_id ObjectId	Title String
1	ObjectId('62bc751dfb5bcaba8d01…	"Cancer prevention: from 1727 …
2	ObjectId('62bc751dfb5bcaba8d01…	"Global Cancer Incidence and M…
3	ObjectId('62bc751dfb5bcaba8d01…	"Measuring cancer evolution fr…
4	ObjectId('62bc751dfb5bcaba8d01…	"Clinical, Prognostic and Ther…
5	ObjectId('62bc751dfb5bcaba8d01…	"Cancer-associated fibroblasts…
6	ObjectId('62bc751dfb5bcaba8d01…	"Tumor microenvironment: recen…
7	ObjectId('62bc751dfb5bcaba8d01…	"What Is Cancer?"
8	ObjectId('62bc751dfb5bcaba8d01…	"Addressing cancer's grand cha…

7.4.3 Consultas Where + Subqueries 3

SQL

```
select DiseaseLabel
from clingene
where Pathways_idPathways1 in
( select idPathways
from pathways
where Subject = 'Metabolic' and Description like '%Ketone%');
```

JSON

```
{{Subject:"Metabolic", Description:/Ketone/}, {DiseaseLabel:1}
```

*Tenemos en cuenta que la colleción de ha creado a partir de un join en Mysql.

Resultados

	_id ObjectId	DiseaseLabel String
1	ObjectId('62bcc24efb5bcaba8d01…	"Baraitser-Winter cerebrofront…

8

Referencias

8.1 Repositorio GitHub

https://github.com/GitHubAlejandroDR/ProyectoFinal_DB.git

UNIVERSIDAD DE MÁLAGA | uma.es

E.T.S de Ingeniería Informática
Bulevar Louis Pasteur, 35
Campus de Teatinos
29071 Málaga

E.T.S. DE INGENIERÍA INFORMÁTICA