

ANEXOS

SISTEMA DE SOPORTE A LA TOMA DE DECISIONES PARA EL PROYECTO DE INVESTIGACIÓN ANÁLISIS MULTIFRACTAL DEL GENOMA HUMANO

ANEXOS

UNIVERSIDAD DEL CAUCA

Alba Viviana Camayo Otero

Adrian Fernando Martinez Molina

**UNIVERSIDAD DEL CAUCA
FACULTAD DE INGENIERÍA ELECTRÓNICA Y TELECOMUNICACIONES
PROGRAMA INGENIERÍA DE SISTEMAS
Grupo de Investigación GTI I + D
Grupo de Investigación y Desarrollo en Tecnologías de la Información
Popayán
2.009**

SISTEMA DE SOPORTE A LA TOMA DE DECISIONES PARA EL PROYECTO DE INVESTIGACIÓN ANÁLISIS MULTIFRACTAL DEL GENOMA HUMANO



UNIVERSIDAD DEL CAUCA

Alba Viviana Camayo Otero

Adrian Fernando Martinez Molina

Trabajo de investigación para optar al título de Ingenieros de Sistemas

Director:

MSc. Martha Eliana Mendoza

**UNIVERSIDAD DEL CAUCA
FACULTAD DE INGENIERÍA ELECTRÓNICA Y TELECOMUNICACIONES
PROGRAMA INGENIERÍA DE SISTEMAS
Grupo de Investigación GTI I + D
Grupo de Investigación y Desarrollo en Tecnologías de la Información
Popayán
2.009**

AGRADECIMIENTOS

Gracias...

A Dios por permitirnos cumplir esta meta y a María Auxiliadora por su protección e intersección.

A nuestros padres: Carmen Lucia Otero, Ángel Miro Camayo, Belén Molina, Servio Martinez y a nuestras familias por su apoyo incondicional, sin su patrocinio y paciencia nada de esto hubiera sido posible.

A nuestra directora, la Magister Martha Eliana Mendoza por su apoyo a lo largo de todo el desarrollo de este proyecto, por su visión y organización, sin esta no habiéramos alcanzado la meta propuesta.

Al Ingeniero Ember Ubeimar Martinez, por sus empecinadas ideas, sin ellas no nos habiéramos enfrentado a tantos desafíos que hicieron de esta experiencia, algo de nunca olvidar.

Al grupo BIMAC, por su apoyo y sus grandes ideas, las cuales son gestoras de más conocimiento, conocimiento que contribuye con el bienestar de la humanidad, en especial a la Profesora Patricia Vélez, gracias por acogernos en su casa (Laboratorio).

A los chicos del laboratorio de Bioinformática, Yohn Jairo Acosta, John Jaime Delgado, Carlos Tellez, Adrian Rodriguez, por los buenos momentos que nos hicieron pasar, por los trasnochos compartidos y por tantos momentos vividos que hacen que de este proyecto una experiencia inolvidable.

A nuestros amigos quienes nos apoyaron a lo largo de la carrera y que nos dieron palabras de ánimo y aliento en todo momento, especialmente a los Joscós, a Rodrigo, Boris y Cuellar, a Diego Ruiz, a Diego Luna, John Betancourt, Hector Alarcon, Jimmy Certuche, Ruben Orozco, Jorge Galindez, Jarvein Mauricio Rivera, Mauricio Hurtado, Gustavo Aponzá, Andres Felipe Manzano, Maria Virginia Lucena, Norma Rivera y Oscar Rendón. Gracias totales incluyendo todos los que se nos olvidan.

Alba Camayo y Adrian Martinez

TABLA DE CONTENIDO

ANEXO A: PREPARACIÓN DE LA ENTREVISTA	1
A.1. INTRODUCCIÓN.....	1
A.2. OBJETIVOS DE LA INVESTIGACIÓN Y CONSIDERACIONES.....	1
ANEXO B: ANÁLISIS DE LOS DATOS.....	4
ANEXO C: CAPTURA DE REQUERIMIENTOS.....	6
C.1. METODOLOGÍAS DE TOMA DE REQUERIMIENTOS.....	6
C.2. ESTUDIO DE LOS DATOS PROPORCIONADOS POR EL GRUPO BIMAC.....	10
C.3. CONSULTAS PRE-ESTABLECIDAS.....	12
ANEXO D: ENTREVISTA REALIZADA EL 5 JUNIO DEL 2008.....	14
ANEXO E: EVALUACIÓN DEL AMBIENTE ORGANIZACIONAL PARA LA EJECUCIÓN DEL PROYECTO.....	22
ANEXO F: PROTOTIPOS DE DATA MART.....	27
ANEXO G: TABLA DE PLANEACIÓN Y SEGIMIENTO.....	30
ANEXO H: PAQUETES DE CARGA ETL.....	31
ANEXO I: ACTION SEQUECES, ESQUEMAS MONDRIAN	37
I.1. DWBIOFINAL.XACTION.....	37
I.2. Dwbiofinal.mondrian.xml.....	40
I.3. Querypivotchr.xaction.....	48
I.4. Consultas mdx.....	52
I.4.1. Análisis FractalAnálisis.....	52
I.4.2. Unidades de información.....	52
I.4.3. Gráfica de cromosomas vs medidas.....	52
ANEXO J: SCRIPT LA BODEGA DE DATOS.....	54
ANEXO K: COMPARACION DE RESULTADO DE CONSULTAS CON DATOS CREADOS POR LOS INVESTIGADORES.	63
ANEXO L: ARQUITECTURA DE HERRAMIENTAS USADAS.....	70
N.1. Arquitectura de Pentaho.....	70
N.2. Arquitectura de Mondrian.....	71
ANEXO M: EJEMPLO DE ARCHIVO GBK.....	72
ANEXO N: PROBLEMAS DE CARGUE DE DATOS BIOLÓGICOS.....	73

N.3. Objetivos de nuestro estudio:.....	73
N.4. Clasificación de los artefactos de los datos.....	74

LISTA DE TABLAS

Tabla C1. Perfiles de Usuario.....	7
Tabla C2. Definición de Roles.....	8
Tabla D1. La voz del cliente.....	13
Tabla D2. Clasificación de Verbalizaciones.....	14
Tabla D3. Estructurar las necesidades del cliente.....	15
Tabla D4. Analizar la estructura de las necesidades del cliente.....	16
Tabla D5. Desglosara las necesidades de alta jerarquía.....	18
Tabla D6. Priorizar las necesidades del cliente.....	19
Tabla E1 .test de litmus Original.....	21
Tabla E2. Test de litmus Adaptado.....	23

LISTA DE FIGURAS

Figura G1.....	31
Figura H1.....	32
Figura H2.....	32
Figura H3.....	33
Figura H4.....	33
Figura H5.....	34
Figura H6.....	34
Figura H7.....	35
Figura H8.....	35
Figura H9.....	36
Figura H10.....	36
Figura H11.....	37
Figura k1.....	64
Figura K2.....	65
Figura K3.....	65
Figura K4.....	66
Figura K5.....	66
Figura K6.....	67
Figura K7.....	67
Figura K8.....	68
Figura K9.....	69
Figura K10.....	69
Figura K11.....	70
Figura K12.....	70
Figura L1.....	71
Figura L2.....	72
Figura M1.....	73
Figura N1.....	76
Figura N2.....	77
Figura N3.....	78
Figura N4.....	79
Figura N5.....	80
Figura N6.....	81
Figura N7.....	82
Figura N8.....	83
Figura N9.....	84
Figura N11.....	86
Figura N12.....	87
Figura N13.....	88
Figura N14.....	89

ANEXO A: PREPARACIÓN DE LA ENTREVISTA

En el presente anexo se describe la preparación que se realizó la primera reunión de facilitación con el grupo de investigación BIMAC, se puede encontrar la definición de roles y preparación de preguntas para el grupo.

A.1. INTRODUCCIÓN

Discuta los objetivos del proyecto de la bodega de datos y su estado general.

El objetivo del proyecto planteado como trabajo de grado es: crear un sistema de soporte a la toma de decisiones para la investigación análisis multifractal del genoma humano; dicho sistema pretende centralizar la información proveniente de las diferentes fuentes de las cuales se extraen los datos que requiere la investigación.

El trabajo de grado persigue generar una herramienta por medio de la cual el grupo de investigadores pueda dar a conocer los resultados de su investigación y además los usuarios de la herramienta puedan manipular los datos y realicen sus propios análisis y saquen sus propias conclusiones.

Hasta el momento se ha logrado explorar en lo que consideramos como un nuevo y diferente escenario de aplicación de las bodegas de datos, esta exploración se ha hecho construyendo una serie de modelos y prototipos con los cuales hemos podido descubrir varios aspectos con los cuales no se contaba inicialmente.

Discuta los objetivos y el curso de la entrevista

El objetivo de esta entrevista es confrontar todos los conceptos y supuestos con los cuales se han venido trabajando, lo que se desea con esta actividad es en parte formalizar el proceso con el grupo de investigadores y de igual manera mantenerlos informados de lo que se ha realizado y lo que falta por realizar.

Presente al equipo de la entrevista y los roles que desempeñan.

Para esta sesión de facilitación contamos con Alba Viviana Camayo quien guiara esta entrevista, Adrian Fernando Martinez quien llevara atenta nota de todo lo que se hable en esta sesión y la Ingeniera Martha Eliana Mendoza quien es la directora del proyecto y el Ingeniero Ember Ubeimar Martinez quien hasta el momento es quien a trabajado conjuntamente con nosotros, colaborándonos en el entendimiento de la problemática, los dos ingenieros desempeñaran el rol de observadores en esta sesión de facilitación.

Describir el paso a seguir después de la entrevista.

El paso a seguir después de esta sesión es, rectificar los modelos que se tienen, depurar los datos con los cuales se ha venido trabajando y avanzar en la construcción de la solución.

A.2. OBJETIVOS DE LA INVESTIGACIÓN Y CONSIDERACIONES

Cuáles son los objetivos de la investigación.

A lo largo de estos meses en los cuales hemos explorado el área en la cual se aplicara la solución planteada en nuestro proyecto de grado creemos que tenemos una visión más clara de lo que se persigue con la investigación.

El objetivo de la investigación visto desde nuestra óptica, es que puedan encontrar coincidencias dentro del genoma de distintos organismos, con el objetivo de realizar una clasificación que tenga

sentido biológico y con la cual puedan contribuir al entendimiento de los genomas de los diferentes organismos que se están estudiando en la investigación.

En cuanto a la herramienta que nos encontramos desarrollando entendemos que contribuirá en la divulgación de los resultados de la investigación, permitiendo que estos resultados sean manipulados por diferentes usuarios que se encuentren navegando en la web y que estén interesados en esta investigación.

Para el desarrollo del proyecto, hemos trabajado hasta el momento con algunos datos con los cuales hemos construido los modelos y realizado pruebas de algunas consultas, las cuales consideramos relevantes evaluar en esta sesión.

Al realizar el proceso para la construcción de los modelos que soportarán estas consultas hemos encontrado algunas dificultades, debido a que las herramientas utilizadas están enfocadas en el contexto de los negocios y se ha dificultado un poco la elaboración de estas consultas, es importante que si existen otras consideraciones o aclaraciones para estas consultas, sean discutidas en este momento.

Iniciaremos la discusión con un conjunto de consultas las cuales manejan un escenario definido por el usuario y un conjunto de datos comunes. El conjunto de datos manejado para estas consultas son: el identificador del gen, identificador de la estructura del gen, la dimensión fractal, el tamaño, el promedio de simetría, el R2 y el número de unidades de información de un gen.

Para la definición del escenario el usuario inicia fijando un límite inferior a partir del cual se seleccionaran todos los R2 superiores a dicho valor y se trabajara con estos valores para realizar los diferentes análisis, posteriormente el usuario establecerá un límite superior, un límite inferior y un valor. Con los límites dados se determinara un intervalo de análisis y el valor determinara el número de rangos en los cuales será dividido el intervalo, los límites de estos rangos son cerrados en su límite superior y abiertos en el inferior. Una vez definido el escenario el usuario podrá observar el número de genes, el promedio de unidades de información y el promedio del porcentaje de simetría, todo de acuerdo a los rangos establecidos y a alguna variable de análisis como por ejemplo la dimensión fractal, la variable de análisis es la que se agrupa de acuerdo a los rangos establecidos.

¿Cuál es el objetivo que persigue esta consulta?

Desearíamos que en este punto se ampliara un poco más acerca de lo que significa el DS y el R2 y cuál es la relación con el resto de información.

Pasando a otra de las consultas teníamos que, dado un número de unidades de información del gen (exones + intrones), el cual será proporcionado por el usuario, se desea obtener todas las estructuras que tengan dicho número de unidades de información, organizar las unidades de información de cada estructura por longitud; se había planteado que esta consulta puede hacerse por cromosoma o por counting, pero el Ingeniero Ember nos comento que es muy posible que el counting no se tenga en cuenta para los análisis.

¿Cuál es el objetivo que persigue esta consulta?

La última consulta que queremos discutir es la de dado un cromosoma o un counting o varios cromosomas o varios counting, obtener todos los genes con su función, Ds y R2.

¿Cuál es el objetivo que persigue esta consulta?

¿Cuál es el objetivo del análisis de las proteínas?

¿Cuáles son los análisis que se hacen sobre las proteínas?

¿Cómo se relacionan las proteínas con el resto de información?

¿Qué tipo de reportes son obtenidos?

Por último quisiéramos discutir acerca de unos ejes centrales de análisis que hemos identificado en la realización de este proyecto el primero es el análisis de la estructura del gen, en este eje de análisis tenemos información acerca de la constitución del gen, sus unidades de información, su funcionalidad y su organización.

Existe un segundo eje de análisis el del análisis multifractal en este eje tenemos información relacionada con las dimensiones fractales, el R2 entre otros.

El último eje de análisis es el de las proteínas este eje aun no ha sido definido y esperamos que con esta sesión podamos determinar la información que se tendrá en cuenta en este eje.

ANEXO B: ANÁLISIS DE LOS DATOS

En el presente documento se estudiarán y se describirán los datos provenientes de la investigación análisis multifractal del genoma humano.

El proceso que sigue el grupo de investigadores para obtener la información necesaria para llevar a cabo la investigación es la siguiente:

Los investigadores acceden a bases de datos encontradas en la red y avaladas por organismos internacionales, los archivos vienen en formato gbk, posteriormente mediante scripts los investigadores analizan la información y generan los datos de interés para la investigación.

Los datos resultantes son almacenados en archivos planos, estos archivos son los que serán utilizados en este proyecto de grado y en base a los cuales se construirá la bodega de datos propuesta, a continuación se mostrarán la descripción de cada uno de los archivos, los datos que contienen, el formato y el tipo de esos datos.

Nombre del archivo: datos.txt

Objetivo:

Datos que contiene:

IDGEN: Identificador del gen

LONGEXON:

LONGINTRON:

LONGEXON:

NUMERO UNIDADES DE INFORMACION:

Nombre del archivo: est.txt

Objetivo:

Datos que contiene:

Range: Rango predefinidos que inician en -0.5 y terminan en el valor dado en la columna.

#Ds: Cantidad de DS que se encuentran en el rango dado.

pSz: Promedio del tamaño del Gen

pUI: Promedio de unidades de información

%Semi: Promedio de simetría

Nombre del archivo: resultado.txt

Objetivo:

Datos que contiene:

IDGEN: Identificador del gen

SIZEGENE: Tamaño del gen

UNIDADESINFORMACION: Suma de los exones e Intrones encontrados en un gen.

DS: 1/pendiente (b1) de la regresión lineal

Coficiente de Determinación (R2):

SIMETRIA: [LENEXON,LENINTRON,LENEXON]

ANEXO C: CAPTURA DE REQUERIMIENTOS

Este anexo contiene la preparación y captura de requerimientos que se llevo a cabo las recomendaciones dadas en el libro de Ralph Kimball y el formato del QFD para la toma de requerimientos. Se da una breve explicación del concepto QFD y de su metodología y a continuación la entrevista y la información recolectada en las plantillas del QFD.

C.1. METODOLOGÍAS DE TOMA DE REQUERIMIENTOS

Ralph Kimball expone en su libro que los usuarios y sus requerimientos causan un gran impacto en las decisiones que han de ser tomadas a largo del proyecto, los requerimientos son el centro del universo de las bodegas de datos, la captura de requerimientos permiten determinar cuales son los datos que deben encontrarse disponibles en la bodega de datos, como deben estar organizados y que tan a menudo esos datos deben ser actualizados; también es importante determinar quienes accederán a la bodega de datos, como se desea que los datos sean presentados y finalmente planear el desarrollo, mantenimiento y crecimiento de la bodega teniendo en cuenta las opiniones del cliente. Se iniciara formulando respuestas a preguntas relacionadas con todo el ciclo de vida y que permita saber que tanto conocimiento tiene a cerca del área en que se planteara la solución.

Es necesario aclarar que en este punto el proyecto ya lleva 5 meses en curso, tiempo en el cual se han realizados algunas actividades en busca de entender y acercarnos más a el ambiente en el cual será implementada la solución, destacando de nuevo que no es el ambiente típico en el que se desempeñan las bodegas de datos, si no por el contrario un ambiente nuevo y con algunas particularidades a tener en cuenta; por tal razón inicialmente se realizaron algunos prototipos con los cuales se buscaba explorar y entender un poco más el ambiente de desarrollo del proyecto, es por esto que en este punto se ha decidido volver a retomar la metodología que se propuso llevar en el desarrollo del proyecto y por medio de la captura de requisitos retroalimentar el proceso y confirmar los supuestos que se han tenido durante los últimos meses.

Para realizar esta actividad y retomando lo planteado por Ralph Kimball, las respuestas a las preguntas planteadas parten de un conocimiento previo en el cual se ha explorado el ambiente por medio de los prototipos creados.

Para guiar el proceso de captura de los requerimientos se hará uso de una plantilla guiada por el concepto del QFD para la captura de requerimientos; el QFD (Despliegue de la función de calidad), es definido por el DR. Yoji Akao, uno de sus desarrolladores; como un método para desarrollar un diseño de calidad dirigido a satisfacer al cliente al traducir sus demandas en metas de diseño y puntos importantes de aseguramiento de la calidad para usarse en toda la fase de producción.

Haremos uso de la metodología del *Blitz QFD*, esta permite alinear nuestros recursos con las verdaderas necesidades del cliente y es una herramienta muy práctica que no requiere de *software* ni de herramientas específicas (como la Casa de la Calidad) para ofrecer resultados (aunque tanto el *software* como la Casa de la Calidad pueden ser también complementos muy útiles al *Blitz QFD*) en la tabla que se muestra a continuación se encuentran los 7 pasos sugeridos por la metodología para la captura y análisis de los requerimientos

Proceso	CAPTURA DE REQUISITOS – QFD						
Propósito	Obtener las necesidades reales del cliente.						
Descripción	<p>Actividades:</p> <p>1. Obtener la Voz del Cliente.</p> <p>Esto implica "ir al lugar de los hechos, ir a donde está la acción"; no se puede escuchar la Voz del Cliente a distancia. Es necesario visitar, preguntar, volver a preguntar y volver a preguntar hasta entender claramente la verbalización de qué es lo que el cliente necesita, para esto se puede grabar lo que el cliente dice y después escribirlo. En esta parte es útil hacer preguntas como:</p> <ul style="list-style-type: none"> - ¿Quién solicitó el sistema? - ¿Quién lo utilizará? - ¿Cuál será el beneficio económico del éxito de la solución? - ¿De qué alternativas se dispone? - ¿Porqué decidieron hacer el sistema? <p>En la mayoría de los casos, cuando el cliente exija una funcionalidad es útil preguntar por qué la necesita o cuál es el objetivo de dicha funcionalidad. Ejemplo:</p> <p>Al preguntar sobre las necesidades del cliente con respecto al diseño de un maletín un cliente comentó: "necesito que sea ligero". Esta necesidad suena obvia en un maletín. Sin embargo, el responsable de la entrevista volvió a preguntar: -¿para qué necesita que sea ligero?- "Para poderlo trasladar fácilmente por el aeropuerto". De aquí podemos obtener una segunda necesidad. Utilicemos una tabla para facilitar la explicación.</p> <table border="1" data-bbox="732 1108 1101 1346"> <thead> <tr> <th data-bbox="738 1117 792 1157">No.</th> <th data-bbox="794 1117 1094 1157">Verbalización</th> </tr> </thead> <tbody> <tr> <td data-bbox="738 1161 792 1241">1</td> <td data-bbox="794 1161 1094 1241">"necesito que sea ligero"</td> </tr> <tr> <td data-bbox="738 1245 792 1341">2</td> <td data-bbox="794 1245 1094 1341">"para poderlo trasladar fácilmente por el aeropuerto"</td> </tr> </tbody> </table> <p>Si el analista en este caso hubiera sólo tomado la primera necesidad explícita, hubiera solicitado a los diseñadores que quitaran todo el peso posible del maletín. Pero al tener una segunda necesidad con una relación jerárquica (necesito A para lograr B), nuestro analista pudo sugerir más opciones, como poner ruedas al maletín para facilitar el transporte.</p> <p>A la hora de entrevistar es muy importante preguntar directamente a la persona que va a utilizar el sistema, ejemplo: secretarias, de nada nos sirve preguntarle al gerente de la empresa como desea llenar un formulario de pago, si al fin de cuentas ese trabajo lo hace la secretaria.</p> <p>2. Clasificar las Verbalizaciones.</p> <p>El objetivo de este paso es clasificar las verbalizaciones por temas afines. Hay "voces del cliente" similares. Otras son complementarias. Otras incluso son opuestas. Para esto lo primero que se debe hacer es establecer un criterio de clasificación: Funcionales, No Funcionales, Técnicas, GUI, Según el tipo de usuario, etc.</p>	No.	Verbalización	1	"necesito que sea ligero"	2	"para poderlo trasladar fácilmente por el aeropuerto"
No.	Verbalización						
1	"necesito que sea ligero"						
2	"para poderlo trasladar fácilmente por el aeropuerto"						

Al clasificar las verbalizaciones, estamos también buscando patrones que nos permitan entender mejor las necesidades del cliente. Algo importante es que éste no es un estudio cuantitativo, sino cualitativo. No nos interesa en esta etapa las estadísticas sobre "el número de verbalizaciones de cada tipo", sino más bien clasificar las verbalizaciones para poder obtener de ellas las necesidades reales del cliente.

3. Estructurar las Necesidades del Cliente.

Una vez que clasificamos las verbalizaciones, tenemos que "extraer" de ellas las necesidades de los clientes. Este es un paso crítico, ya que algunas son explícitas y muy claras; otras son implícitas y algunas hasta nos podrán parecer absurdas. Sin embargo, es vital recordar que estamos buscando las necesidades reales del cliente, no "nuestra versión de las necesidades del cliente"; hay que cuidarnos de nuestro propio sesgo.

No.	Verbalización	Necesidad Detectada
1	"necesito que sea ligero"	Necesito que sea ligero
2	"para poderlo trasladar fácilmente por el aeropuerto"	Necesito que sea fácil de trasladar en lugares donde hay mucho tráfico de personas

4. Analizar la Estructura de las Necesidades del Cliente.

Hay necesidades que tienen relaciones de dependencia. Por ejemplo, en el caso anterior Notemos como el poner ruedas resolvería la Necesidad 2, pero va en contra de la Necesidad 1. En al Captura de requisitos, nos interesan las necesidades de más alta jerarquía, ya que son éstas las que más impacto (positivo o negativo), tienen sobre nuestros clientes.

5. Desglosar las necesidades de Alta jerarquía (nivel 1)

Para esto se usará un diagrama de árbol, el cual es una herramienta analítica que nos sirve para establecer relaciones causales entre diferentes necesidades de cliente, una vez que logramos extraer estas necesidades de la voz del cliente.

Una vez que logramos extraer las necesidades de la voz del cliente, con esta técnica las podemos agrupar en Necesidades Nivel 1 (más generales) y Necesidades Nivel 2, 3...,n (que son más específicas). Se recomienda tener 3 Niveles de Necesidades máximo.

De acuerdo con Mazur [1] típicamente habrán entre 5-10 Necesidades Nivel 1 que se dividirán en 20-40 Necesidades Nivel 2 y 40-80 Necesidades Nivel 3... Se recomienda buscar 7 ± 2 Necesidades Nivel 1 (5 a 9). Estas son las más generales y serán comparables entre sí desde el punto de vista del cliente.

Se ponen estas necesidades en el lado izquierdo de la página, para hacer un árbol horizontal (son generalmente más fáciles). Por ejemplo, tomemos las necesidades de un grupo de clientes con respecto a la bebida (soda, refresco) que se sirve en un restaurante.

	<p><i>Ejemplo de Necesidades de Nivel 1</i></p> <p><i>Bebida sabrosa</i> <i>Vaso limpio</i> <i>Servicio rápido</i></p> <p>El árbol se va generando a través de preguntar ¿cómo cumplir? a cada una de estas necesidades... De cada necesidad Nivel 1, deben salir 3±2 necesidades de Nivel 2 (entre 2 y 5)...</p> <p><i>Ejemplo de Necesidades de Nivel 2</i></p> <p><i>Bebida sabrosa</i> ---- <i>Bebida fría</i> ---- <i>-Bebida en buen estado</i></p> <p>Las necesidades de Nivel 3 salen de las Necesidades de Nivel 2... Se deben tener nuevamente 3±2 necesidades de Nivel 3 para cada Necesidad de Nivel 2... Hay proyectos que no son tan complicados y no ameritan más que 1 ó 2 niveles; por otra parte, entre más niveles se tengan, es necesario hacer un mayor esfuerzo de análisis, pero el beneficio es que se tendrá un estudio más completo...</p> <p><i>Ejemplo de Necesidades de Nivel 3</i></p> <p><i>Bebida sabrosa</i> --- <i>Bebida fría</i> ---- <i>Vaso frío</i> ---- <i>Vaso con hielos</i></p> <p>6. Priorizar las Necesidades del Cliente. Esto implica establecer cuáles necesidades de primer nivel son más importantes para nuestros clientes. ¿Bueno, Bonito o Barato? Si le diéramos a nuestro cliente \$100 para invertirlo en necesidades ¿cuánto nos compraría de cada una? ¿\$50 en Bueno, \$25 en Bonito y \$25 en Barato? La mejor forma de hacer esto, es una vez identificadas las necesidades y estratificadas, preguntar directamente a los clientes.</p> <p>7. Analizar sólo las relaciones prioritarias a detalle. Al evaluar nuestro producto o servicio, los puntos más importantes son aquellos que impactan a las necesidades prioritarias. Si le vamos a pedir al cliente que nos evalúe, hay que enfocarse en aquellos elementos que impactan a las necesidades prioritarias. Aquí es donde debemos enfocar nuestros recursos, ya que el nivel de calidad de nuestros productos y servicios estará determinado por la medida en que logremos alinear el valor de los recursos con la prioridad de las necesidades de nuestros clientes.</p>
--	--

Tabla C1. Perfiles de Usuario.

Ahora definiremos de manera clara las actividades y roles que se requieran para llevar a cabo la actividad planteada.

Para la captura de los requerimientos se realizará una sesión de facilitación con el grupo BIMAC, en esta actividad se presentara el último prototipo construido y se plantearan algunas preguntas,

esta actividad tiene la intención de proveer retroalimentación al proceso y confirmar o refutar los supuestos sobre los que se ha venido construyendo el sistema.

DEFINICIÓN DE ROLES

Para la sesión de facilitación se solicitara a los usuarios que pueda ser usada una grabadora con el fin de capturar de la mejor forma todos los requerimientos y sugerencias del cliente, en la tabla que se presentara a continuación se presentarán los roles que se requieren para la actividad.

Rol	Descripción	Quien lo desempeña
Líder de la entrevista	Es la persona encargada de dirigir la actividad, realizar las preguntas y también tomar nota de las cosas que considere importantes	ALBA VIVIANA CAMAYO
Escriba	Es el encargado de tomar atenta nota durante la entrevista, el escriba captura con mayor detalle las apreciaciones de los participantes de la actividad.	ADRIAN FERNANDO MARTINEZ
Observadores	Como su mismo nombre lo dice su papel es observar.	MARTHA ELIANA MENDOZA, EMBER UBEIMAR MARTINEZ.

Tabla C2. Definición de Roles.

C.2. ESTUDIO DE LOS DATOS PROPORCIONADOS POR EL GRUPO BIMAC.

A continuación se estudiarán y se describirán los datos provenientes de la investigación análisis multifractal del genoma humano.

El proceso que sigue el grupo de investigadores para obtener la información necesaria para llevar a cabo la investigación es la siguiente:

Los investigadores acceden a bases de datos encontradas en la red y avaladas por organismos internacionales, los archivos vienen en formato gbk, posteriormente mediante scripts los investigadores analizan la información y generan los datos de interés para la investigación.

Los datos resultantes son almacenados en archivos planos, estos archivos son los que serán utilizados en este proyecto de grado y en base a los cuales se construirá la bodega de datos propuesta, a continuación se mostraran la descripción de cada uno de los archivos, los datos que contienen, el formato y el tipo de esos datos.

Nombre del archivo: resultado.txt

Objetivo:

Datos que contiene:

IDGEN: Identificador del gen

SIZEGENE: Tamaño del gen

UNIDADESINFORMACION: Suma de los exones e Intrones encontrados en un gen.

DF: 1/pendiente (b1) de la regresión lineal

Coeficiente de Determinación (R2):

SIMETRIA: [LENEXON, LENINTRON, LENEXON]

Nombre del archivo: gen.txt

Objetivo:

Datos que contiene:

IDCHR: Identificador del cromosoma.

IDGEN: Identificador del gen.

INICIO: Posición de inicio del gen dentro del counting.

FIN: Posición de fin del gen dentro del counting.

LONGITUD: Tamaño del gen.

BANDA: En que hélice se encuentra.

IDCOUNTING: Identificador del counting.

Nombre del archivo: BodegasCHR_NN.txt

Objetivo:

Datos que contiene:

IDCHR: Identificador del cromosoma.

IDGEN: Identificador del gen.

IDCOUNTING: Identificador del counting.

IDESTRUCTURA: Identificador de la estructura.

DESCRIPCIÓN: Nombre de la unidad de información.

ORDEN TIPO: Orden en que se encuentran organizadas las unidades de información con la misma descripción dentro del gen.

ORDEN ESTRUCTURA: Orden en que se encuentran organizadas las unidades de información dentro del gen.

INICIO: Posición de inicio de la unidad de información dentro del counting.

FIN: Posición de fin de la unidad de información dentro del counting.

LONGITUD: Tamaño de la unidad de información.

VERSIÓN DEL COUNTING: Versión del counting.

C.3. CONSULTAS PRE-ESTABLECIDAS.

- Dado un valor determinado por el usuario obtener todos los R2 mayores a dicho valor y:
 - Determinar el número de genes con un DS, que se encuentra en cualquiera de los rangos determinados; dichos rangos son definidos a partir de una distancia la cual es definida por el usuario.
 - Promedio de longitudes de genes con un DS, que se encuentra en cualquiera de los rangos determinados; dichos rangos son definidos a partir de una distancia la cual es definida por el usuario.
 - Promedio de las unidades de información de los genes con un DS, que se encuentra en cualquiera de los rangos determinados; dichos rangos son definidos a partir de una distancia la cual es definida por el usuario.
 - Porcentaje de simetría de los genes con un DS, que se encuentra en cualquiera de los rangos determinados; dichos rangos son definidos a partir de una distancia la cual es definida por el usuario.
- Obtener todas las estructuras que tienen un determinado número de unidades de información, definido por el usuario y organizar por longitud. Esta consulta puede hacerse para un cromosoma o un counting o varios cromosomas o varios counting.
- Dado un cromosoma o un counting o varios cromosomas o varios counting, obtener todos los genes con su función, Ds y R2.

Definición de las preguntas de la sesión de facilitación

Aclarar:

- Objetivo de la investigación.
- Las consultas pre-establecidas.
- El alcance del proyecto.
- La visualización de la información (datos que serán visualizados)
- Definición de los términos.
- Definición de los usuarios finales.

Introducción

Objetivos De La investigación Análisis Multifractal Del Genoma Humano

General:

Modelar el genoma de algunos organismos eucariotes y en especial el genoma humano con herramientas derivadas del análisis multifractal con el propósito de generar una clasificación con sentido biológico y de gran utilidad en el estudio de la salud humana y su biotecnologización.

Específicos:

- Cuantificar el grado de variación genética del cromosoma eucariote (según el modelo) en función de la anatomía molecular conocida del cromosoma.
- Generar una clasificación de las secciones del genoma que tengan un sentido biológico y que permita su uso biotecnológico (en el análisis estructural y funcional del genoma).

Objetivos De La Construcción del DSS.

Objetivo General

Desarrollar un sistema de soporte a la toma de decisiones estratégicas (DSS) basado en tecnologías de DW y OLAP, que centralice los datos provenientes de las diferentes fuentes de la investigación “Análisis Multifractal del Genoma Humano para la Búsqueda de Regularidades con Significado Biológico y una Contribución a la Generación de Biotecnología de la Información”, permitiendo el análisis que de esta información deben realizar los investigadores.

Objetivos Específicos

- Diseñar y construir un DW que centralice los datos provenientes de la investigación “Análisis Multifractal del Genoma Humano” y del GenBank, permitiendo almacenar información relacionada con los genes, unidades de información y proteínas de dicha investigación.
- Construir un prototipo de herramienta OLAP que utilice herramientas y/o componentes, que permita representar de forma gráfica y en grillas los datos obtenidos de la investigación “Análisis Multifractal del Genoma Humano” para su respectivo análisis.

Generar recomendaciones para el diseño y construcción de un DW en el campo de la bioinformática, que aporten a la construcción de aplicaciones que tengan características similares a la investigación “Análisis Multifractal del Genoma Humano”, buscando de esta manera contribuir en el área de Aplicaciones de un DW.

ANEXO D: ENTREVISTA REALIZADA EL 5 JUNIO DEL 2008

En este anexo de contiene las plantilladas del método QFD para la captura de requisitos que se llevo a cabo en la entrevista que se realizo con el investigador líder del grupo BIMAC.

Actividad 1: Obtener La Voz del Cliente	
Preguntas	Respuestas
¿Cuál es el objetivo de la investigación Análisis Multifractal del Genoma Humano?	Buscar las regularidades existentes en las Bases de Datos de los genomas, particularmente en el genoma humano. Una vez se han encontrado dichas regularidades, se busca un significado biológico a las mismas.
¿Quién lo utilizará?	Los integrantes del grupo BIMAC, investigadores interesados en el tema.
¿Por Qué decidieron hacer el sistema?	El sistema contribuirá con la divulgación de los resultados obtenidos en la investigación, además de permitir que los usuarios extraigan mas información manipulando los datos almacenados.
Observaciones: Para la captura de la vos del cliente se realizo una entrevista (se grabó).	
Verbalizaciones	
1	“Necesitamos analizar una variable, la cual se dividirá en rangos y se le aplicara una serie de filtros, para el análisis requerido”.
2	“Necesitamos que los intervalos y rangos de análisis sean configurables”.
3	“Se tienen una plantillas, las cuales proporcionan unos límites, rangos y otros valores preestablecidos.”
4	“Necesitamos que el usuario tenga la opción de configurar otros límites, rangos u otros valores ”
5	“Necesitamos que el usuario tenga la opción de filtrar los datos obtenidos de los rangos, dando un valor mínimo de R ² ”
6	“Necesitamos que una vez los datos se encuentren filtrados, el usuario pueda visualizar el número de genes que se encuentran en cualquiera de los rangos de análisis.”
7	“Necesitamos que una vez los datos se encuentren filtrados, el usuario pueda visualizar el promedio de unidades de información que se encuentran en cualquiera de los rangos de análisis.”
8	“Necesitamos que una vez los datos se encuentren filtrados, el usuario pueda visualizar el promedio de longitudes de los genes que se encuentran en cualquiera de los rangos de análisis.”
9	“Necesitamos que una vez los datos se encuentren filtrados, el usuario pueda visualizar el promedio de los genes que se encuentran en cualquiera de los rangos de análisis.”
10	“Necesitamos que el usuario visualice gráficamente la relación entre el numero de genes y los rangos definidos.”
11	“Necesitamos que el usuario visualice gráficamente la relación entre el promedio de genes y los rangos definidos.”
12	“Necesitamos que el usuario visualice gráficamente la relación entre el promedio de unidades de información y los rangos definidos.”
13	“Necesitamos que el usuario visualice gráficamente la relación entre el promedio de longitudes de los genes y los rangos definidos.”
14	“Necesitamos que el usuario visualice gráficamente una tabla en la que se encuentre por

	cada rango definido la siguiente información: .”
15	“Necesitamos que el usuario pueda acceder a cada cuadro de la tabla anterior, para que conozca al detalle la información que se encuentra en ese rango.”
16	“Necesitamos que el usuario pueda definir un numero de unidades de información dentro de un cromosoma determinado, extrayendo todas las configuraciones que tengan el mismo número de unidades de información determinado por el usuario, una vez obtenidas las configuraciones, se deben organizar de acuerdo a la longitud de las unidades de información.”
17	“Necesitamos obtener todos los genes con su función, R2 y D, de acuerdo a un cromosoma dado”

Tabla D1. La voz del cliente.

Actividad 2: Clasificar las Verbalizaciones	
Criterio Clasificación: Según el tipo de análisis.	
Categoría 1: <Análisis Fractal>	
Verbalizaciones	
1	“Necesitamos analizar una variable, la cual se dividirá en rangos y se le aplicara una serie de filtros, para el análisis requerido”
2	“Necesitamos que los intervalos y rangos de análisis sean configurables”
3	“Se tienen una plantillas, las cuales proporcionan unos límites, rangos y otros valores preestablecidos.”
4	“Necesitamos que el usuario tenga la opción de configurar otros límites, rangos u otros valores ”
5	“Necesitamos que el usuario tenga la opción de filtrar los datos obtenidos de los rangos, dando un valor mínimo de R2.”
6	“Necesitamos que una vez los datos se encuentren filtrados, el usuario pueda visualizar el número de genes que se encuentran en cualquiera de los rangos de análisis.”
7	“Necesitamos que una vez los datos se encuentren filtrados, el usuario pueda visualizar el promedio de unidades de información que se encuentran en cualquiera de los rangos de análisis.”
8	“Necesitamos que una vez los datos se encuentren filtrados, el usuario pueda visualizar el promedio de longitudes de los genes que se encuentran en cualquiera de los rangos de análisis.”
9	“Necesitamos que una vez los datos se encuentren filtrados, el usuario pueda visualizar el promedio de los genes que se encuentran en cualquiera de los rangos de análisis.”
10	“Necesitamos que el usuario visualice gráficamente la relación entre el numero de genes y los rangos definidos.”
11	“Necesitamos que el usuario visualice gráficamente la relación entre el promedio de genes y los rangos definidos.”
12	“Necesitamos que el usuario visualice gráficamente la relación entre el promedio de unidades de información y los rangos definidos.”
13	“Necesitamos que el usuario visualice gráficamente la relación entre el promedio de longitudes de los genes y los rangos definidos.”
14	“Necesitamos que el usuario visualice gráficamente una tabla en la que se encuentre por cada rango definido la siguiente información: .”
15	“Necesitamos que el usuario pueda acceder a cada cuadro de la tabla anterior, para que conozca al detalle la información que se encuentra en ese rango.”
15	“Necesitamos obtener todos los genes con su función, R2 y D, de acuerdo a un cromosoma dado”

Categoría 2: <Análisis de Unidades de Información> Verbalizaciones	
1	“Necesitamos que el usuario pueda definir un número de unidades de información dentro de un cromosoma determinado, extrayendo todas las configuraciones que tengan el mismo número de unidades de información determinado por el usuario, una vez obtenidas las configuraciones, se deben organizar de acuerdo a la longitud de las unidades de información.”

Tabla D2. Clasificación de Verbalizaciones.

Actividad 3: Estructurar las Necesidades del Cliente.			
Categoría 1: <Análisis Fractal>			
	Verbalizaciones	Necesidades Detectadas	Código
1	“Necesitamos analizar una variable, la cual se clasificara en rangos y se le aplicara una serie de filtros, para el análisis requerido”	Necesitamos Clasificar en rangos los datos proporcionados.	AF-N1
2	“Necesitamos que los intervalos y rangos de análisis sean configurables”	Necesitamos generar intervalos y rangos de análisis.	AF-N2
3	“Se tienen una plantillas, las cuales proporcionan unos límites, rangos y otros valores preestablecidos.”	Necesitamos cargar todos los datos conforme a los rangos preestablecidos	AF-N3
4	“Necesitamos que el usuario tenga la opción de configurar otros límites, rangos u otros valores ”	Necesitamos generar consultas dinámicas.	AF-N4
5	“Necesitamos que el usuario tenga la opción de filtrar los datos obtenidos de los rangos, dando un valor mínimo de R2.”	Necesitamos filtrar los datos obtenidos de los rangos de análisis, dando un valor mínimo con el cual se obtendrán los datos que tengan un R2 superior al valor mínimo dado.	AF-N5
6	“Necesitamos que una vez los datos se encuentren filtrados, el usuario pueda visualizar el número de genes que se encuentran en cualquiera de los rangos de análisis.”	Necesitamos que el usuario visualice los datos obtenidos de la consulta que le permite ver la cantidad de genes que se encuentra en cualquiera de los rangos de análisis.	AF-N6
7	“Necesitamos que una vez los datos se encuentren filtrados, el usuario pueda visualizar el promedio de unidades de información que se encuentran en cualquiera de los rangos de análisis.”	Necesitamos que el usuario visualice los datos obtenidos de la consulta que le permite ver el promedio de unidades de información se encuentra en cualquiera de los rangos de análisis.	AF-N7
8	“Necesitamos que una vez los datos se encuentren filtrados, el usuario pueda visualizar el promedio de longitudes de los genes que se encuentran en cualquiera de los rangos de análisis.”	Necesitamos que el usuario visualice los datos obtenidos de la consulta que le permite ver el promedio de las longitudes de los genes que se encuentra en cualquiera de los rangos de análisis.	AF-N8
9	“Necesitamos que una vez los datos se encuentren filtrados, el usuario pueda visualizar el promedio de los genes que se encuentran en	Necesitamos que el usuario visualice los datos obtenidos de la consulta que le permite ver el promedio de los genes que se encuentra en cualquiera	AF-N9

	cualquiera de los rangos de análisis.”	de los rangos de análisis.	
10	“Necesitamos que el usuario visualice gráficamente la relación entre el número de genes y los rangos definidos.”	Necesitamos graficar la relación entre el número de genes y los rangos definidos.	AF-N10
11	“Necesitamos que el usuario visualice gráficamente la relación entre el promedio de genes y los rangos definidos.”	Necesitamos graficar la relación entre el promedio de genes y los rangos definidos.	AF-N11
12	“Necesitamos que el usuario visualice gráficamente la relación entre el promedio de unidades de información y los rangos definidos.”	Necesitamos graficar la relación entre el promedio de unidades de información y los rangos definidos.	AF-N12
13	“Necesitamos que el usuario visualice gráficamente la relación entre el promedio de longitudes de los genes y los rangos definidos.”	Necesitamos graficar la relación entre el promedio de longitudes de los genes y los rangos definidos.	AF-N13
14	“Necesitamos que el usuario visualice gráficamente una tabla en la que se encuentre por cada rango definido la siguiente información: .”	Necesitamos construir una tabla dinámica, en la cual el usuario pueda visualizar los datos clasificados dentro de los rangos.	AF-N14
15	“Necesitamos que el usuario pueda acceder a cada cuadro de la tabla anterior, para que conozca al detalle la información que se encuentra en ese rango.”	Necesitamos que el usuario pueda obtener información mas detallada de la tabla de clasificación generada.	AF-N15
16	“Necesitamos obtener todos los genes con su función, R2 y D, de acuerdo a un cromosoma dado”	Necesitamos obtener todos los genes con su función, R2 y D de acuerdo a un cromosoma dado	AF-N16
Categoría 2: <Análisis de las Unidades de Información>			
	Verbalizaciones	Necesidades Detectadas	Código
1	“Necesitamos que el usuario pueda definir un número de unidades de información dentro de un cromosoma determinado, extrayendo todas las configuraciones que tengan el mismo número de unidades de información determinado por el usuario, una vez obtenidas las configuraciones, se deben organizar de acuerdo a la longitud de las unidades de información.”	Necesitamos una consulta en la cual el usuario pueda definir un número de unidades de información y a partir de esto le arroje todas las estructuras de un cromosoma dado que tienen dicha cantidad de unidades de información para posteriormente organizar las estructuras de acuerdo a la longitud de sus unidades de información.	AUI-N1

Tabla D3. Estructurar las necesidades del cliente.

Actividad 4: Analizar la Estructura de las Necesidades del Cliente.		
Categoría 1: <Análisis Fractal>		
Código	Necesidad	Código de las Necesidades de las cuales Depende
AF-N1	Necesitamos Clasificar en rangos los datos proporcionados.	-

AF-N2	Necesitamos generar intervalos y rangos de análisis.	
AF-N3	Necesitamos cargar todos los datos conforme a los rangos preestablecidos.	AF-N1, AF-N2
AF-N4	Necesitamos generar consultas dinámicas.	
AF-N5	Necesitamos filtrar los datos obtenidos de los rangos de análisis, dando un valor mínimo con el cual se obtendrán los datos que tengan un R2 superior al valor mínimo dado.	AF-N1, AF-N2, AF-N3, AF-N4
AF-N6	Necesitamos que el usuario visualice los datos obtenidos de la consulta que le permite ver la cantidad de genes que se encuentra en cualquiera de los rangos de análisis.	AF-N1, AF-N2, AF-N3, AF-N4
AF-N7	Necesitamos que el usuario visualice los datos obtenidos de la consulta que le permite ver el promedio de unidades de información se encuentra en cualquiera de los rangos de análisis.	AF-N1, AF-N2, AF-N3, AF-N4
AF-N8	Necesitamos que el usuario visualice los datos obtenidos de la consulta que le permite ver el promedio de las longitudes de los genes que se encuentra en cualquiera de los rangos de análisis.	AF-N1, AF-N2, AF-N3, AF-N4
AF-N9	Necesitamos que el usuario visualice los datos obtenidos de la consulta que le permite ver el promedio de los genes que se encuentra en cualquiera de los rangos de análisis.	AF-N1, AF-N2, AF-N3, AF-N4
AF-N10	Necesitamos graficar la relación entre el número de genes y los rangos definidos.	AF-N1, AF-N2, AF-N3, AF-N4
AF-N11	Necesitamos graficar la relación entre el promedio de genes y los rangos definidos.	AF-N1, AF-N2, AF-N3, AF-N4
AF-N12	Necesitamos graficar la relación entre el promedio de unidades de información y los rangos definidos.	AF-N1, AF-N2, AF-N3, AF-N4
AF-N13	Necesitamos graficar la relación entre el promedio de longitudes de los genes y los rangos definidos.	AF-N1, AF-N2, AF-N3, AF-N4
AF-N14	Necesitamos construir una tabla dinámica, en la cual el usuario pueda visualizar los datos clasificados dentro de los rangos.	AF-N1, AF-N2, AF-N3, AF-N4, AF-N5, AF-N6, AF-N7, AF-N8
AF-N15	Necesitamos que el usuario pueda obtener información más detallada de la tabla de clasificación generada.	AF-N1, AF-N2, AF-N3, AF-N4, AF-N5, AF-N6, AF-N7, AF-N8
AF-N16	Necesitamos obtener todos los genes con su función, R2 y D de acuerdo a un cromosoma dado	
Categoría 2: <Análisis de Unidades de Información>		
Código	Necesidad	Código de las Necesidades de las cuales Depende
AUI-N1	Necesitamos una consulta en la cual el usuario pueda definir un número de unidades de información y a partir de esto le arroje todas las estructuras de un cromosoma dado que tienen dicha cantidad de unidades de información para posteriormente organizar las estructuras de acuerdo a la longitud de sus unidades de información.	

Tabla D4. Analizar la estructura de las necesidades del cliente

Actividad 5: Desglosar las necesidades de Alta jerarquía			
Categoría 1: < Análisis Fractal >			
Código	Necesidades		
	Nivel 1	Nivel 2	Nivel 3
AF-N1	Necesitamos Clasificar en rangos los datos proporcionados.	El usuario determina el intervalo de análisis el cual puede estar preestablecido ó definido por el usuario.	El usuario determina en cuantos rangos se dividirá el intervalo de análisis.
			El usuario puede determinar el tamaño de los rangos.
AF-N5	Necesitamos filtrar los datos obtenidos de los rangos de análisis, dando un valor mínimo con el cual se obtendrán los datos que tengan un R2 superior al valor mínimo dado.	El usuario selecciona un valor preestablecido.	Obtener la cantidad de genes que se encuentra en cualquiera de los rangos de análisis.
			Obtener el promedio de unidades de información se encuentra en cualquiera de los rangos de análisis.
			Obtener el promedio de las longitudes de los genes que se encuentra en cualquiera de los rangos de análisis.
		El usuario determina un valor diferente a los preestablecidos.	Graficar la relación entre el número de genes y los rangos definidos.
			Graficar la relación entre el promedio de unidades de información y los rangos definidos.
			Graficar la relación entre el promedio de longitudes de los genes y los rangos definidos.
AF-N14	Necesitamos construir una tabla dinámica, en la cual el usuario pueda visualizar los datos clasificados dentro de los rangos.	El usuario selecciona los datos que desea visualizar de acuerdo a un intervalo y un rango dado	Visualización de la consulta en la herramienta OLAP
AF-N15	Necesitamos que el usuario pueda obtener información mas detallada de la tabla de clasificación generada.	El usuario realiza actividades de navegación en la OLAP	Establecer tareas típicas de la herramienta OLAP.

AF-N16	Necesitamos obtener todos los genes con su función, R2 y D de acuerdo a un cromosoma dado	El usuario adiciona las dimensiones y medidas que requiera a la herramienta OLAP	Permitir la adición de dimensiones y medidas en la herramienta OLAP
Categoría 2: < Análisis de Unidades de Información>			
Código	Necesidad		
	Nivel 1	Nivel 2	Nivel 3
CL-N1	Necesitamos una consulta en la cual el usuario pueda definir un número de unidades de información y a partir de esto le arroje todas las estructuras de un cromosoma dado que tienen dicha cantidad de unidades de información para posteriormente organizar las estructuras de acuerdo a la longitud de sus unidades de información.	El usuario determina el número de unidades de información.	Permitir al usuario determinar el número de unidades de información que desea analizar
		El usuario determina el cromosoma a ser analizado.	Permitir al usuario escoger el cromosoma que desea analizar
		El sistema debe arrojar todas las estructuras que cumplan con las restricciones puestas por el usuario y organizarlas de acuerdo a la longitud de cada una de las unidades de información	Visualizar el resultado de la consulta según las especificaciones del usuario

Tabla D5. Desglosara las necesidades de alta jerarquía

Actividad 6: Priorizar las Necesidades del Cliente.			
#	Código	Necesidad	Prioridad (Alta - Media - Baja)
1	AF-N1	El usuario determina en cuantos rangos se dividirá el intervalo de análisis.	ALTA
2	AF-N1	El usuario puede determinar el tamaño de los rangos.	ALTA
3	AF-N5	Obtener la cantidad de genes que se encuentra en cualquiera de los rangos de análisis.	ALTA
4	AF-N5	Obtener el promedio de unidades de información se encuentra en cualquiera de los rangos de análisis.	ALTA
5	AF-N5	Obtener el promedio de las longitudes de los genes que se encuentra en cualquiera de los rangos de análisis.	ALTA
6	AF-N5	Graficar la relación entre el número de genes y los rangos definidos	ALTA
7	AF-N5	Graficar la relación entre el promedio de unidades de información y los rangos definidos.	ALTA
8	AF-N5	Graficar la relación entre el promedio de longitudes de los genes y los rangos definidos.	ALTA
9	AF-N14	Visualización de la consulta en la herramienta OLAP	ALTA
	CL-N1	Permitir al usuario determinar el número de unidades de información que desea analizar	ALTA
	CL-N1	Permitir al usuario escoger el cromosoma que desea analizar	

	CL-N1	Visualizar el resultado de la consulta según las especificaciones del usuario	ALTA
1	AF-N15	Establecer tareas típicas de la herramienta OLAP.	MEDIA
2	AF-N16	Permitir la adición de dimensiones y medidas en la herramienta OLAP	MEDIA

Tabla D6. Priorizar las necesidades del cliente

ANEXO E: EVALUACIÓN DEL AMBIENTE ORGANIZACIONAL PARA LA EJECUCIÓN DEL PROYECTO.

Este anexo expone una de las tareas propuestas por Ralph Kimball, en su libro del ciclo de vida dimensional,

Aquí se encuentran algunos factores que permiten evaluar la disposición que una organización tiene para el desarrollo de un DW y de esta manera establecer las bases para asegurar el éxito de un proyecto DW; sin embargo, estos factores no aplican ampliamente en el área en la que este proyecto fue desarrollado, no obstante se analizaron estos factores y se adaptaron al área de aplicación, como se mostrara a continuación.

A continuación en la tabla 1 se presentara el test de Litmus, realizando posteriormente y de acuerdo a los resultados obtenidos, un análisis de los cinco factores evaluados, con el fin de tener una visión clara de la situación de la organización respecto al proyecto.

Factor	Baja Disposición	⇔	Alta Disposición
Fuerte Patrocinio de la Gestión Empresarial			
	Con no muy buena reputación	⇔	Con considerable influencia en la organización
	Puede llevar semanas al equipo acceder a el	⇔	Disponible para el equipo
	“Voy a volver a usted en que”	⇔	Rápido, decisivo en la resolución de los asuntos
	Tiene la esperanza de que “usted” lo logre	⇔	Activo, elocuente y es un soporte – dispuesto a poner su propia cabeza en juego
	¿Usted puede entregar esto a 250 usuarios el próximo mes, verdad?	⇔	Expectativas realistas
	“¿Qué es un DW?”	⇔	Con conocimientos en las DW
Obligada por una Motivación Empresarial			
	“¿Y tu punto es?”	⇔	La supervivencia depende de la DW
	La financiación es un gran problema	⇔	Los costos no son un problema – nosotros no podemos permitirnos no hacerlo!
	Visión como “arenas movedizas”	⇔	Visión claramente articulada
	Diez diferentes puntos de vista de la solución	⇔	Consistente panorama de la solución
	Cuestiones tácticas	⇔	Cuestiones estratégicas
	Oportunidad de ahorro de costos	⇔	Oportunidad de incrementar los ingresos

	No se puede cuantificar el retorno de la inversión	⇔	Gran retorno de la inversión
Sociedad entre los Representantes del Negocio/IS			
	El negocio contrata un consultor externo sin conocimiento del IS	⇔	El negocio y el IS trabajan conjuntamente
	EL negocio crea una unidad propia con un supuesto equipo del IS para construir una DW	⇔	El IS comprometidos activamente con las unidades del negocio t
	"No podemos confiar en cualquiera de nuestros sistemas"	⇔	Fuerte confianza en los reportes del entorno
	Se tarda "años" en tener una nueva solicitud ad hoc de vuelta	⇔	Rápida respuesta del IS a las solicitudes ad hoc
	Los usuarios ni siquiera envían mas solicitudes	⇔	Periodos cortos entre las solicitudes de los usuarios
Cultura Analítica (Actualidad)			
	"Se siente vacío" con la toma de decisiones	⇔	La toma de decisiones basado en hechos y cifras
	Los usuarios no piden datos	⇔	Los usuarios del negocio claman por el acceso a los datos - "solo denme los datos y yo obtengo cifras"
	Los usuarios no se identifican con los reportes actuales	⇔	Los reportes actuales son regularmente transcritos a hojas de calculo para el análisis y la tendencia histórica
	Los reportes actuales son usados como "parador de puertas" hasta que se convierte en papelera de reciclaje	⇔	Los reportes actuales son como los oídos de un sabueso, de importancia y lleno de colores y notas autoadhesivas
	Los usuarios tienen secretarías que imprimen y leen su correo electrónico	⇔	Los usuarios son literalmente muy computarizados
	Las fianzas son extremadamente bien cuidadas del desempeño de las Finance is extremely possessive of bottom line performance figures	⇔	La información es compartida abiertamente a toda la información
Viabilidad			
	La DW requiere la compra de toda la nueva tecnología	⇔	Robusta infraestructura técnica
	Todo el mundo esta comprometido hasta el año 3000 con el proyecto	⇔	Experimentando con los recursos disponibles
	Los datos fiables no estarán disponibles hasta la implementación del Sistema de planificación de recursos de la empresa (ERP)	⇔	Datos de calidad disponibles

Tabla E1 .test de litmus Original

Factor	Ambiente con disponibilidad baja		Ambiente con disponibilidad alta	
Influencia del sponsor en la organización	No muy respetado		Tiene una influencia considerable en la organización	X
	Puede tardar semanas en acceder al equipo		Tiene fácil acceso al equipo	X
	“Me contactare con usted después”		Solución rápida y decisiva de los problemas	X
	Espero que lo hayas hecho		Es una persona activa, dispuesta a interceder por la iniciativa.	X
	Expectativas poco realistas frente al proyecto	X	Expectativas realistas frente al proyecto	
	Tiene poco conocimiento acerca de las bodegas de datos	X	Tiene un amplio conocimiento en el tema de las bodegas de datos	
Motivación del proyecto	¿Y el punto es?	X	La supervivencia de las investigaciones depende de la bodega de datos	
	La financiación del proyecto representa un gran problema		El costo no es un problema	X
	No se tiene una visión clara de lo que se quiere		Se tiene una visión articulada	X
	Se tiene múltiples vistas de la solución		Se tiene una visión unificada de lo que se quiere	X
	Cuestiones tácticas		Cuestiones estratégicas	X
	No se puede cuantificar el retorno de la inversión	X	Se tiene una alto retorno de la inversión	
Relación entre equipos de la organización (Área Biológica – Área tecnologías de la información)	Los investigadores y el grupo de tecnologías de la información trabajan por separado.		Los investigadores y el grupo de tecnologías de la información trabajan conjuntamente.	X
	El equipo del área biológica crea su propio equipo de pseudo- grupo de tecnologías de la información para construir la bodega de datos		El equipo del área biológica colabora activamente con el equipo de tecnologías de la información.	X
	No consideran conveniente mostrar los resultados de sus investigaciones.		Existe bastante confianza en la existencia de un entorno de presentación de reportes.	X
	Las solicitudes realizadas tardan mucho en ser respondidas.		Las solicitudes que el equipo de tecnologías de la información presenta son atendidas rápidamente.	X

Presencia de cultura analítica	Confían en la intuición para poder tomar decisiones.		La toma de decisiones se basa en datos y cifras.	X
	Los datos no tienen mayor relevancia para analizar diferentes aspectos del equipo.		Por medio de los datos se conoce el estado de los procesos y los avances de la investigación.	X
	Los informes de la investigación no son revisados meticulosamente.		Los informes son constantemente organizados en hojas de cálculo para el análisis de tendencias.	X
	Los informes se tienen como papel de reciclaje.		Los informes son constantemente consultados.	X
	Los investigadores no manejan programas computacionales que permitan realizar análisis.		Los investigadores son muy diestros en el manejo de programas que le permitan realizar análisis sobre sus datos.	X
	El manejo de los datos es restringido.		La organización comparte los datos sin muchas restricciones.	X
Factibilidad	La construcción de la bodega de datos requiere de la compra de nuevas tecnologías.		La organización cuenta con una infraestructura técnica robusta.	X
	No existen recursos disponibles.		Existen recursos disponibles.	X
	Los datos no se encuentran disponibles en el momento.		Los datos están disponibles y son de calidad.	X

Tabla E2. Test de litmus Adaptado

Es importante destacar que se omitieron algunos ítems del test al igual que fueron modificados otros, debido al carácter del proyecto; recordemos que las bodegas de datos están orientadas al sector de negocios por esta razón la bibliografía consultada contienen una terminología orientada a tal fin, se considera necesario entonces depurar la información, ajustar los términos, los modelos al desarrollo de este proyecto específico.

A continuación se presentaran cada uno de los cinco factores evaluados con su correspondiente análisis, resultado del test realizado para tal fin.

Influencia del sponsor en la organización: El proyecto se desarrollara como un complemento de la investigación análisis multifractal del genoma humano, investigación que esta siendo desarrollada por el grupo BIMAC de la Universidad del Cauca y los cuales son parte del centro de investigaciones de bioinformática a nivel nacional, dentro del equipo de investigadores se encuentra el Ingeniero Ember Ubeimar Martínez Flor, el se denominara como sponsor, es la persona con la cual se ha tenido mas contacto, sin negar el respaldo del resto del grupo y adicionar que se cuenta con un equipo computacional dentro del laboratorio de los investigadores, manteniendo contacto con todo el equipo.

En conclusión se dirá que el sponsor es de gran influencia en la organización, debido a que pertenece al equipo de investigadores, además se cuenta con todo el apoyo del resto del equipo.

Motivación del proyecto: La investigación análisis multifractal realizada por el grupo BIMAC de la universidad del Cauca ha generado en los últimos años un gran volumen de información relativa al genoma humano, la investigación busca encontrar patrones dentro del genoma humano y esperan que a la culminación de la investigación la información pueda ser manipulada por usuarios a quienes les interese dicha información y que puedan realizar diferentes análisis sobre los datos presentados.

La bodega de datos es una alternativa de solución, con la cual se pretende generar un sistema en el que se encuentren los datos integrados

Acompañamiento del grupo de investigadores: El grupo de investigadores se encuentran muy motivados con el proyecto y han proporcionado las herramientas y los espacios requeridos para el desarrollo del proyecto.

Presencia de cultura analítica: El grupo BIMAC, es un grupo dedicado a realizar estudios relacionados con el entorno biológico, el análisis es una labor que se desempeña a diario.

Factibilidad: Debido a que el proyecto se encuentra enmarcado dentro de un trabajo de grado de la universidad del Cauca, se cuenta con los recursos que proporciona la Universidad y el grupo de investigaciones, además se cuenta con disponibilidad de los datos, los cuales son proporcionados por el grupo de investigación.

Como se puede apreciar de 25 ítems con los que se cuentan para evaluar los cinco factores, cuatro ítems son de disponibilidad baja, lo que nos lleva a concluir que se tienen un ambiente de disponibilidad alta para la realización del proyecto.

ANEXO F: PROTOTIPOS DE DATA MART

En este anexo se muestran los primeros prototipos de data mart que fueron creados, puesto que el proyecto se enfrentaba al modelado de datos biológicos y la documentación existente era escasa.

F1. Data mart Análisis Fractal

Dimensión Dim_Chr:

En esta dimensión están almacenados los cromosomas de un organismo determinado. Esta dimensión contiene:

- **ChrID:** Identificador del cromosoma.
- **Description:** Descripción del cromosoma.
- **Division:**
- **Organism:** Organismo al que pertenece el genoma

Dimensión Dim_Estruct:

En esta dimensión se almacenan la o las estructuras que forman a cada Gen. Esta dimensión contiene:

- **EstructID:** Identificador de la Estructura.
- **D:** Dimensión fractal calculada a la estructura de un Gen.
- **R2:** Regresión lineal.

Dimensión Dim_Gen:

En esta dimensión están almacenados todos los genes existentes dentro de cada cromosoma. Esta dimensión contiene:

- **GenID:** Identificador del Gen.
- **Band:** Banda ó hélice del ADN en la que se encuentra ubicado el gen.
- **Family:** Familia a la cual pertenece el Gen.
- **Begin:** Comienzo del Gen.
- **End:** Fin del Gen.

Tabla de hechos Fact_Analisis_fractal:

Medidas:

- **#UI:** Número de unidades de información existentes en un Gen específico.
- **Simetria:** Porcentaje de simetría de un Gen específico.
- **Size:** Longitud de Gen (en pares base pb).

Consulta Requerida:

Orden de la consulta

1. Seleccionando un organismo.
2. De acuerdo a una distancia de D (de 0 – 1 ó de 0 – 2).
3. Seleccionar una cantidad de rangos para dividir dicha distancia.
4. Se puede seleccionar todos los R2 ó filtrarlos de acuerdo a las necesidades del usuario si este necesita R2's mayores a un valor específico.
5. Seleccionar los Genes con un número de unidades de información específica.

6. Seleccionar los Genes con un tamaño específico.
7. Seleccionar los genes pertenecientes a una familia.
8. Número de Genes.
9. Promedio del Tamaño de los Genes.
10. Promedio de porcentaje de simetría de los Genes, promedio del número de unidades de información.

F2. Data mart Unidades de información

Dimensión Dim_Chr:

En esta dimensión están almacenados los cromosomas de un organismo determinado. Esta dimensión contiene:

- **ChrID:** Identificador del cromosoma.
- **Description:** Descripción del cromosoma.
- **Division:**
- **Organism:** Organismo al que pertenece el genoma

Dimensión Dim_Estruct:

En esta dimensión se almacenan la o las estructuras que forman a cada Gen. Esta dimensión contiene:

- **EstructID:** Identificador de la Estructura.
- **D:** Dimensión fractal calculada a la estructura de un Gen.
- **R2:** Regresión lineal.

Dimensión Dim_Gen:

En esta dimensión están almacenados todos los genes existentes dentro de cada cromosoma. Esta dimensión contiene:

- **GenID:** Identificador del Gen.
- **Band:** Banda ó hélice del ADN en la que se encuentra ubicado el gen.
- **Family:** Familia a la cual pertenece el Gen.
- **Begin:** Comienzo del Gen.
- **End:** Fin del Gen.

Dimensión Dim_Tipo:

En esta dimensión están almacenados todas las unidades de información (exones e Intrones) existentes dentro de cada estructura. Esta dimensión contiene:

- **Tipoid:** Identificador de la unidad de información.
- **Descripción:** tipo de la unidad de información.
- **Begin:** Comienzo de la unidad de información.
- **End:** Fin de la unidad de información.

Tabla de hechos Fact_Analisis_unidades de información:

Medidas:

- **Orden tipo:** Orden de las unidades de información respecto a su tipo si son exones o Intrones.
- **begin:** inicio de la unidad de información

- **end:** final de unidad de información
- **Size:** Longitud de la unidad de información (en pares base pb).

ANEXO G: TABLA DE PLANEACIÓN Y SEGUIMIENTO

Este anexo presenta la tabla de planeación y seguimiento del proyecto.

Actividad	Sub – Actividad	Cod	Responsable (s)	Esfuerzo Estimado	Fecha de inicio	Fecha fin estimada	Fecha Fin	Estado	observaciones	Esfuerzo	Dependencias	Días de Retraso
Plan del Proyecto	Definición de Actividades				02/05/2008	02/05/2007		f				
	Definición de Roles				02/05/2008	03/05/2008		f				
	Plan de comunicaciones				02/05/2008	01/10/2008		f				
	Plan de riesgos				06/10/2008	10/10/2008		f				
Definición de Requerimientos (QFD)	Preparar las entrevistas o la sesión				03/05/2008	06/05/2008		f				
	Obtener la voz del cliente				06/05/2008	06/05/2008		f				
	Clasificar las verbalizaciones				06/05/2008	11/05/2008		f				
	Estructurar las necesidades del				06/05/2008	11/06/2008		f				
	Analizar la estructura de las				06/05/2008	12/06/2008		f				
	Desglosar las necesidades de alta				06/05/2008	13/06/2008		f				
	Priorizar las necesidades del cliente				14/06/2008	06/18/2008		f				
RUTA DE LOS DATOS												
Modelo dimensional	Matriz Bus				21/07/2007	25/07/2007		f				
	Diagrama de Tablas de Hechos				28/07/2007	15/08/2007		f				
	Diagrama detallado de cada Tabla de				04/08/2007	15/08/2007		f				
	identificar los orígenes de datos de				18/08/2007	29/08/2007		f				
	mapa de origen y destino de los				25/08/2007	29/08/2007		f				
Diseño físico	realizar diseño de agregación.				01/09/2007	05/09/2007		f				
								f				
Carga de datos, diseño y desarrollo, ETL	primer cargue (prototipo)				02/10/2007	17/10/2007	30/10/2007	f	datos con inconsistencias editar los archivos			-13
	segundo cargue				04/04/2008	18/04/2008	07/07/2008	f	datos con inconsistencias editar los archivos			-80
	cargue final (familias subfamilias descripciones)							f	datos con inconsistencias se solicitaron los archivo originales			-51
					21/02/2009	25/02/2009	#####					
RUTA TECNOLÓGICA												
Diseño de la arquitectura					10/09/2007	19/11/2007		f				
Selección e instalación de					10/09/2007	19/11/2007	06/06/2008	f	sin documentacion, ni soporte			-200
RUTA DE APLICACIÓN												
Especificación de las					10/09/2007	19/11/2007		f				
Aplicaciones de desarrollo					10/09/2007	19/11/2007		f				
Despliegue					23/04/2009	03/05/2009		f				
Mantenimiento y crecimiento								f				
Administración								f				

Figura G1 tabla de Planeación y seguimiento.

ANEXO H: PAQUETES DE CARGA ETL

Este el anexo contiene pantallazos de todos los paquetes de ETL que se construyeron para poblar la DW con los datos del grupo BIMAC

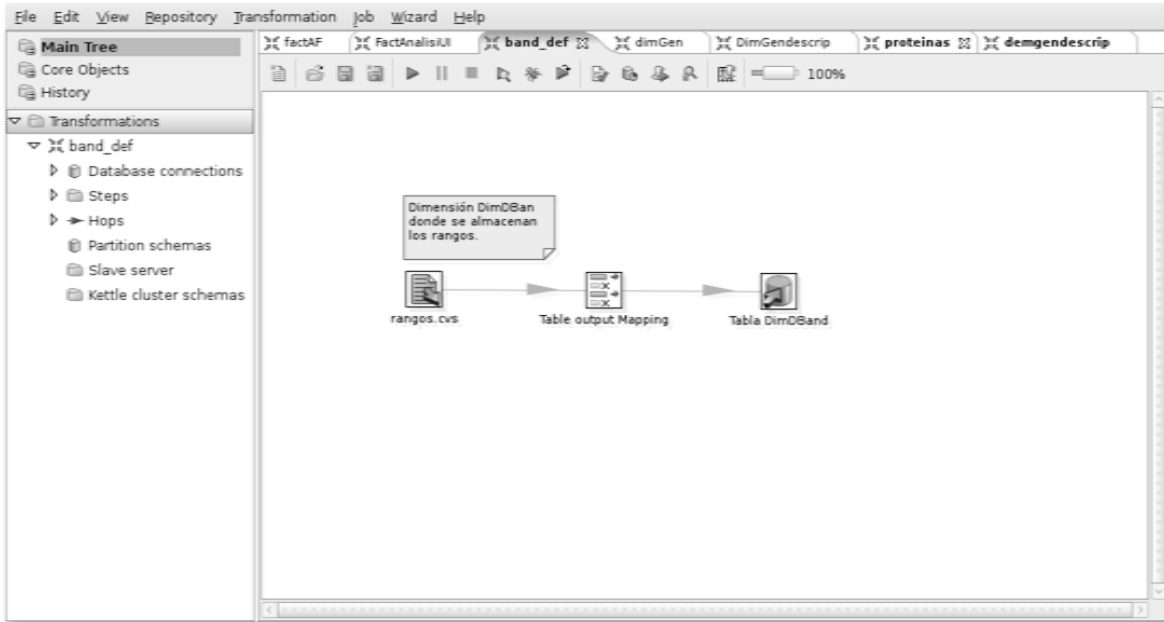


Figura H1. Band_def

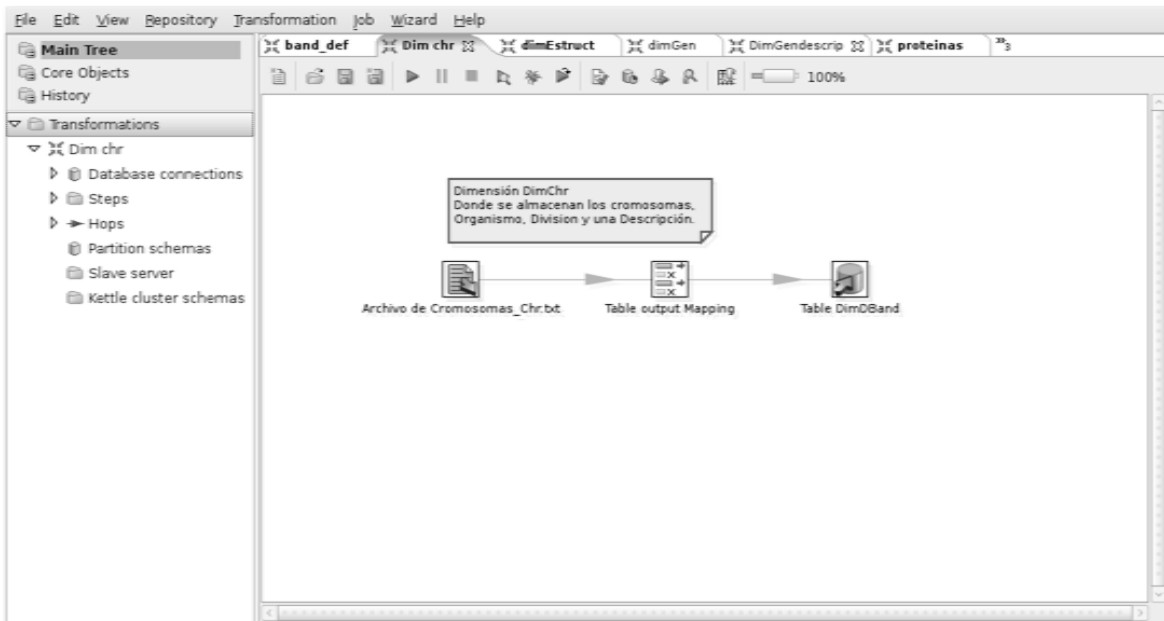


Figura H2. Dim_Chrcr

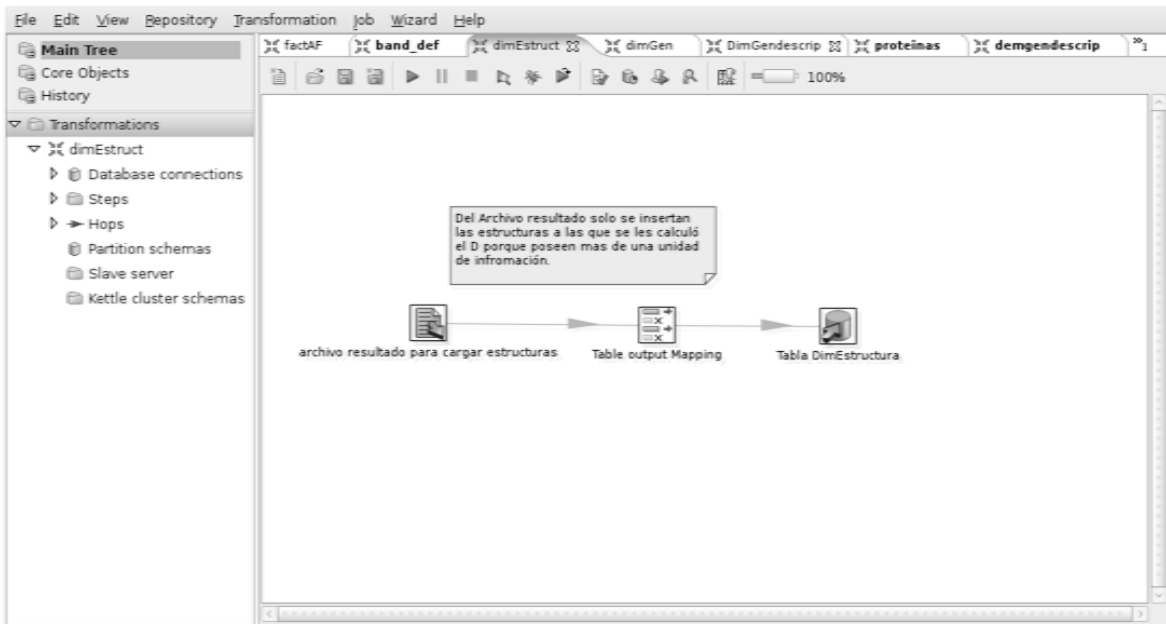


Figura H3. Dim_estruct

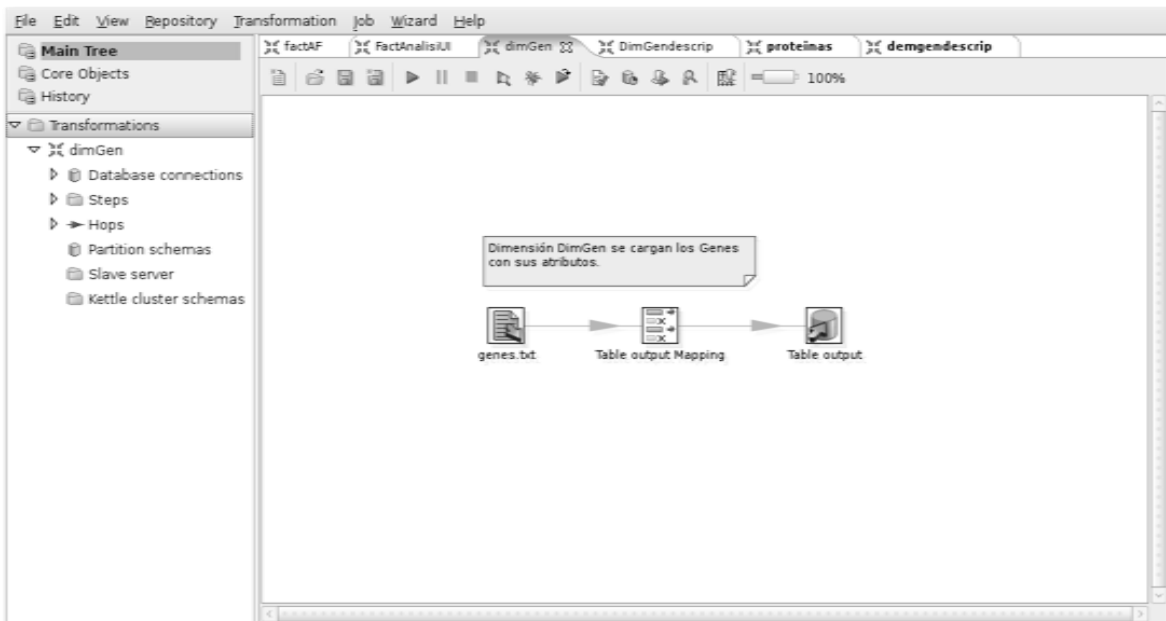


Figura H4. Dim_gen

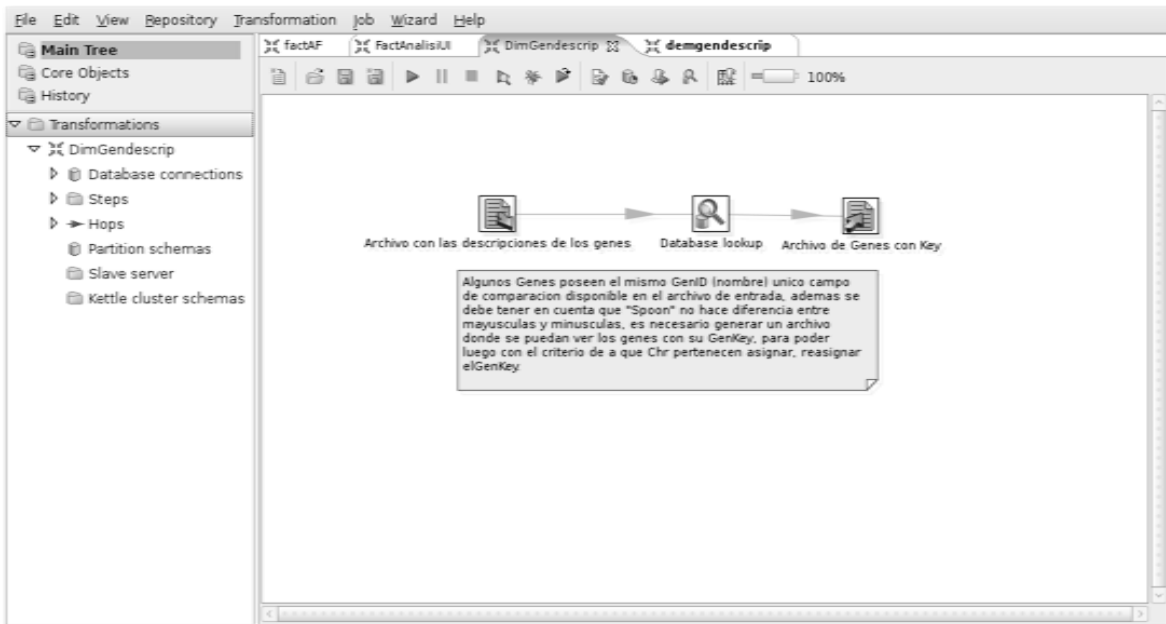


Figura H5. DimGenDescript

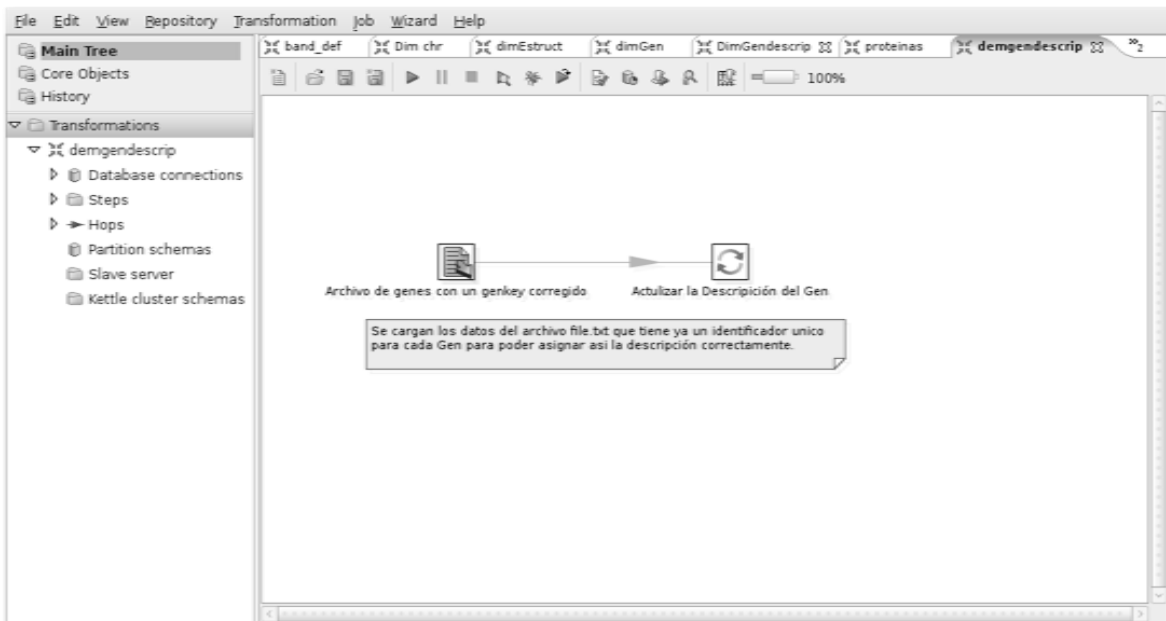


Figura H6. DimGenDescript2

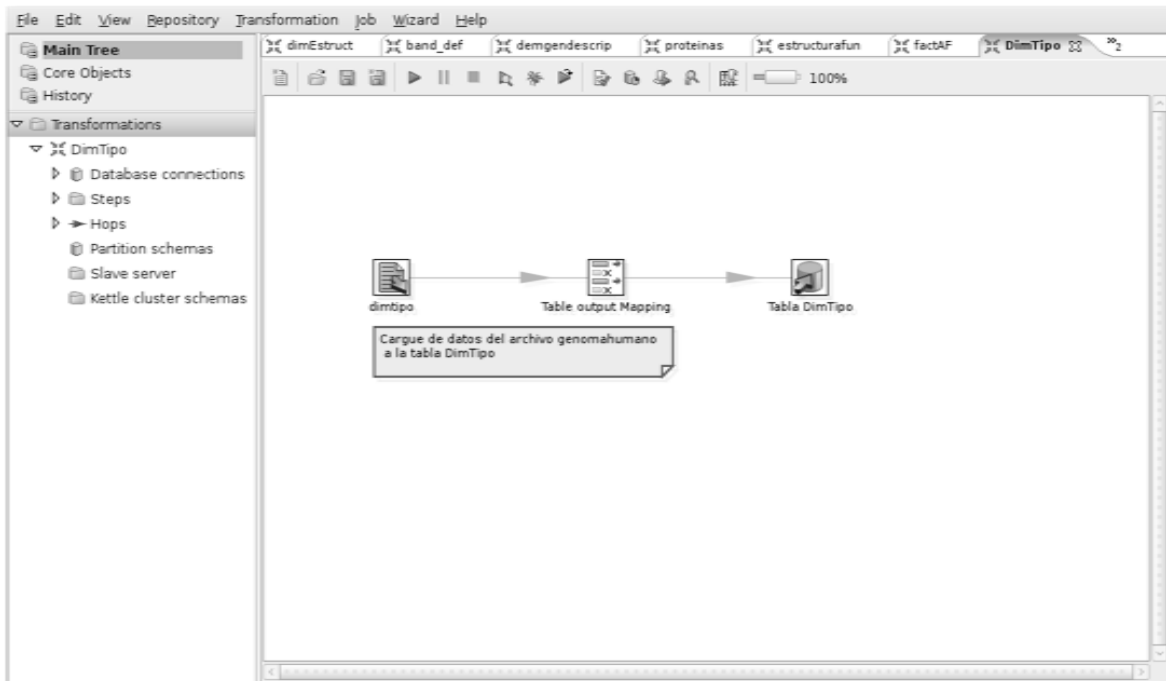


Figura H7. DimTipo

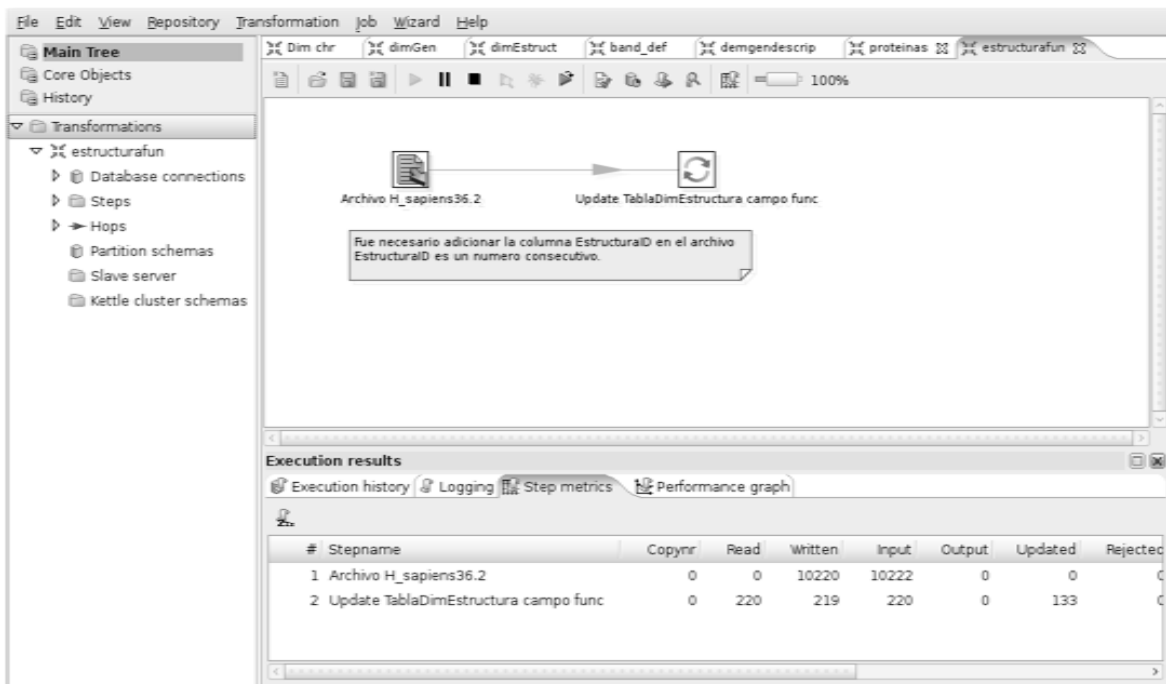


Figura H8. Estructura funcion



Figura H9. Proteinas, Familias, Subfamilias.

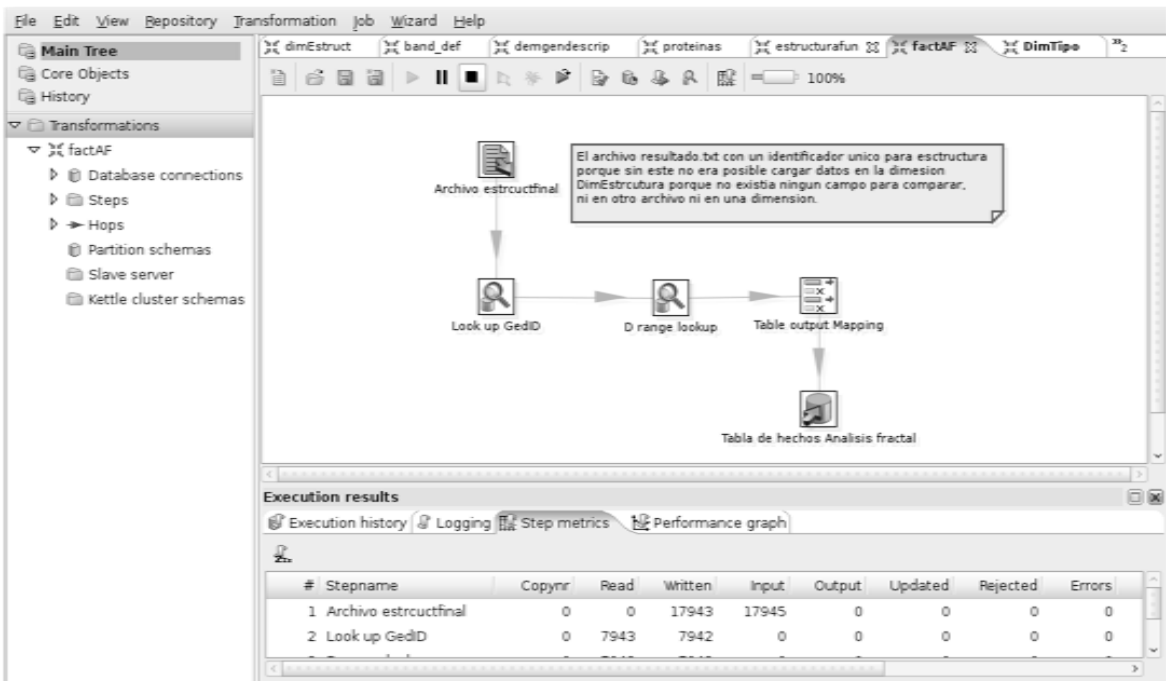


Figura H10. Tabla de Hechos Analisis Fractal

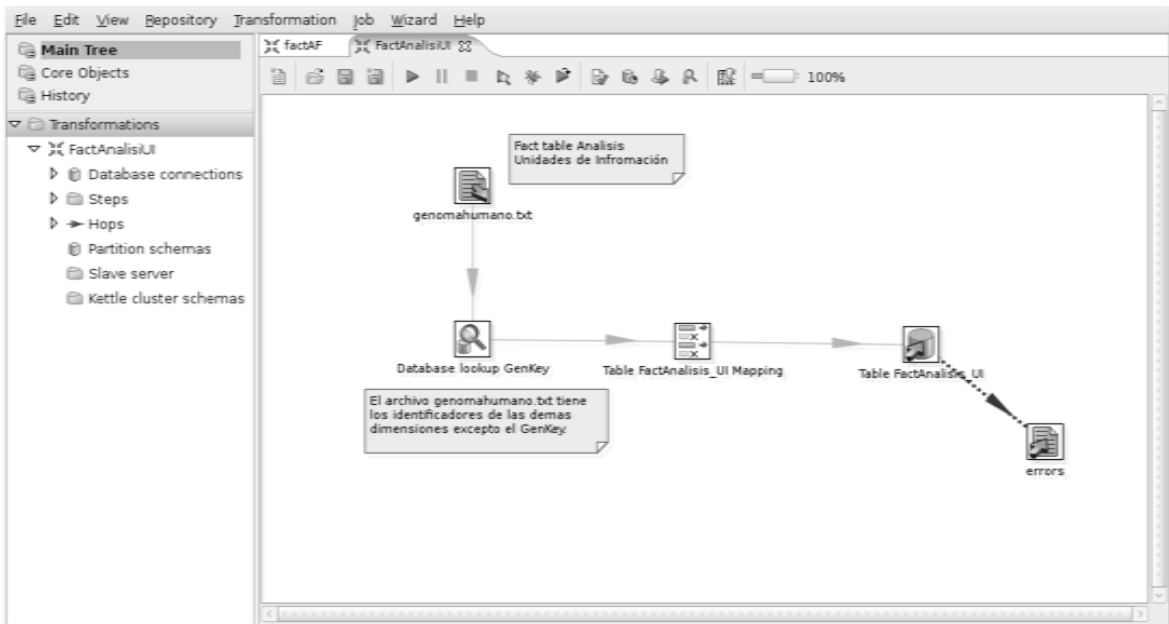


Figura H11. Tabla de hechos analisis unidades de informacion.

ANEXO I: ACTION SEQUECES, ESQUEMAS MONDRIAN

Este Anexo muestra los Actions sequences que fueron creados, para el desarrollo de esta solución en Pentaho, por la poca documentación gratuita existente de esta herramienta es de gran importancia para trabajos futuros contar con estos actions sequence que hacen uso de gran parte de la funcionalidad de Pentaho.

I.1. DWBIOFINAL.XACTION

Este action sequence es utilizado para el primer data mart Análisis de unidades de información, crea una vista en Jpivot con el esquema estrella creado en mondrian como solución al DW.

```
<?xml version="1.0" encoding="UTF-8"?>
<action-sequence>
  <name>DwBiofinal.xaction</name>
  <version>1</version>
  <title>Análisis de Unidades de Información</title>
  <documentation>
    <author>GTI + Bimac</author>
    <description>%description</description>
    <help>To edit this action-sequence use the Pentaho Design Studio.</help>
    <result-type>report</result-type>
    <icon>slicedice.png</icon>
  </documentation>
  <inputs>
    <mode type="string">
      <default-value/>
      <sources>
        <request>mode</request>
      </sources>
    </mode>
  </inputs>
  <outputs>
    <model type="string"/>
    <connection type="string"/>
    <mdx type="string"/>
    <options type="string"/>
    <title type="string"/>
    <url type="string">
      <destinations>
        <response>redirect</response>
      </destinations>
    </url>
    <chartaxistickfontsize type="string"/><chartslicerfontsize
type="string"/><chartlegendfontfamily type="string"/><showgrid type="string"/><actionreference
type="string"/><chartaxislabelfontsize type="string"/><chartslicerlocation
type="string"/><chartaxistickfontfamily type="string"/><chartheight
type="string"/><chartslicerfontfamily type="string"/><chartverticalaxislabel
```

```
type="string"/><chartaxislabelfontstyle type="string"/><chartdrillthroughenabled
type="string"/><chartbackgroundb type="string"/><chartwidth type="string"/><role
type="string"/><chartaxistickfontstyle type="string"/><charttitlefontsize
type="string"/><chartbackgroundg type="string"/><chartlegendfontstyle
type="string"/><chartshowslicer type="string"/><charttitlefontstyle
type="string"/><chartlegendlocation type="string"/><chartlocation
type="string"/><chartbackgroundr type="string"/><charttype
type="string"/><chartaxislabelfontfamily type="string"/><chartlegendfontsize
type="string"/><chartshowlegend type="string"/><chartaxisticklabelrotation
type="string"/><charthorizaxislabel type="string"/><charttitle
type="string"/><chartsliceralignment type="string"/><chartslicerfontstyle
type="string"/><charttitlefontfamily type="string"/></outputs>
<resources/>
<actions>
  <action-definition>
    <action-name>Pivot View</action-name>
    <action-inputs>
      <mode type="string"/>
    </action-inputs>
    <action-outputs>
      <model type="string"/>
      <connection type="string"/>
      <mdx type="string"/>
      <options type="list"/>
      <title type="string"/>
      <url type="string"/>
      <chartaxistickfontsize type="string"/><chartslicerfontsize
type="string"/><chartlegendfontfamily type="string"/><showgrid type="string"/><actionreference
type="string"/><chartaxislabelfontsize type="string"/><chartslicerlocation
type="string"/><chartheight
type="string"/><chartverticalaxislabel
type="string"/><chartdrillthroughenabled
type="string"/><chartwidth type="string"/><role
type="string"/><charttitlefontsize
type="string"/><chartlegendfontstyle
type="string"/><charttitlefontstyle
type="string"/><chartlocation
type="string"/><charttype
type="string"/><chartlegendfontsize
type="string"/><chartaxisticklabelrotation
type="string"/><charttitle
type="string"/><chartslicerfontstyle
type="string"/><charttitlefontfamily type="string"/></action-outputs>
    <component-name>PivotViewComponent</component-name>
    <component-definition>
      <title>Análisis de Unidades de Información</title>
      <viewer>Pivot</viewer>
```

```
<model>solution:BIMAC/datawarehouse/DwBiofinal.mondrian.xml</model>
<jndi>dwbiofinal</jndi>
<query>select NON EMPTY {[Measures].[Size], [Measures].[Orden_tipo],
[Measures].[Pos_final], [Measures].[Pos_inicial]} ON COLUMNS,
NON EMPTY Order({([DimGen.genid].[genes], [estructs.numui].[All ui],
[DimTipo.DimTipo].[uis])}, [Measures].[Size], DESC) ON ROWS
from [dwbio_ui]
where [DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr21]
</query>
<options>
  <personal/>
  <cube-nav/>
  <mdx-edit/>
  <sort-conf/>
  <level-style/>
  <hide-spans/>
  <properties/>
  <non-empty/>
  <swap-axes/>
  <drill-member/>
  <drill-position/>
  <drill-replace/>
  <drill-thru/>
  <chart/>
  <chart-conf/>
  <print-conf/>
  <print-pdf/>
  <excel/>
</options>
<chartaxistickfontsize>10</chartaxistickfontsize><chartslicerfontsize>12</chartslicerfontsize><
chartlegendfontfamily>SansSerif</chartlegendfontfamily><showgrid>true</showgrid><actionref
erence>BIMAC/datawarehouse/DwBiofinal.xaction</actionreference><chartaxislabelfontsize>1
2</chartaxislabelfontsize><chartslicerlocation>1</chartslicerlocation><chartaxistickfontfamily>S
ansSerif</chartaxistickfontfamily><charheight>300</charheight><chartslicerfontfamily>SansS
erif</chartslicerfontfamily><chartvertaxislabel></chartvertaxislabel><chartaxislabelfontstyle>0</
chartaxislabelfontstyle><chartdrillthroughenabled>false</chartdrillthroughenabled><chartbackg
roundb>255</chartbackgroundb><chartwidth>1200</chartwidth><role/><chartaxistickfontstyle>
0</chartaxistickfontstyle><charttitlefontsize>18</charttitlefontsize><chartbackgroundg>255</ch
artbackgroundg><mdx>select NON EMPTY {[Measures].[Size], [Measures].[Orden_tipo],
[Measures].[Pos_final], [Measures].[Pos_inicial]} ON COLUMNS,
NON EMPTY Order({([DimGen.genid].[genes], [estructs.numui].[All ui],
[DimTipo.DimTipo].[uis])}, [Measures].[Size], DESC) ON ROWS
from [dwbio_ui]
where [DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr21]
</mdx><chartlegendfontstyle>0</chartlegendfontstyle><chartshowslicer>true</chartshowslicer
><charttitlefontstyle>1</charttitlefontstyle><connection/><chartlegendlocation>3</chartlegendlo
cation><chartlocation>none</chartlocation><chartbackgroundr>255</chartbackgroundr><chartt
```

```

ype>1</charttype><chartaxislabelfontfamily>SansSerif</chartaxislabelfontfamily><chartlegendfont
onsize>10</chartlegendfontsize><chartshowlegend>true</chartshowlegend><chartaxisticklabel
elrotation>90</chartaxisticklabelrotation><charthorizaxislabel></charthorizaxislabel><charttitle>
</charttitle><chartsliceralignment>3</chartsliceralignment><chartslicerfontstyle>0</chartslicerf
ontstyle><charttitlefontfamily>SansSerif</charttitlefontfamily></component-definition>
    </action-definition>
</actions>
</action-sequence>
    
```

1.2. Dwbiofinal.mondrian.xml

Esquema Dimensional de la DW creado en WorkBench, es un archivo XML en el cual se definen todas las dimensiones, tablas de hecho, medidas y medidas calculadas (que en nuestro caso no eran de utilidad), para la construcción de esta DW se crearon en el esquema dos cubos, uno por cada data mart, para el data mart de análisis de unidades de información se creó el cubo dwbioui, para el data mart Análisis fractal se creó dwbiof, adicionalmente durante la exploración de posibles soluciones a los requerimientos de los investigadores se creó un cubo virtual (fui) que tenía por dimensiones todas aquellas dimensiones compartidas entre el cubo dwbiof y dwbioui, y sirve para poder tener acceso a las medidas que comparten los cubos sin tener en cuenta la granularidad, sin embargo no es posible acceder a las dimensiones no compartidas.

```

<Schema name="dwbiof">
    <Dimension name="DimChr">
        <Hierarchy name="Dimchr" hasAll="true" allMemberName="AllOrganisms"
        primaryKey="ChrKey" primaryKeyTable="DimChr">
            <Table name="DimChr">
                </Table>
            <Level name="organismo" column="Organismo" type="String"
            uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
                </Level>
            <Level name="chrId" table="DimChr" column="ChrID" type="String"
            uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
                </Level>
        </Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" name="DimGen">
        <Hierarchy name="genid" hasAll="true" allMemberName="genes"
        primaryKey="GenKey" primaryKeyTable="DimGen">
            <Table name="DimGen">
                </Table>
            <Level name="genid" table="DimGen" column="GenID" type="String"
            uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
                </Level>
            <Level name="dercip" table="DimGen" column="descrip" type="String"
            uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
                </Level>
        </Hierarchy>
    
```



```
<Hierarchy name="Familia" hasAll="true" allMemberName="Familias"
primaryKey="GenKey" primaryKeyTable="DimGen">
  <Table name="DimGen">
  </Table>
  <Level name="Familia" table="DimGen" column="familia" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
  </Level>
  <Level name="subfamilia" table="DimGen" column="subfamilia"
type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
  </Level>
  <Level name="genid" table="DimGen" column="GenID" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
  </Level>
  <Level name="descripcion" table="DimGen" column="descrip"
type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
  </Level>
</Hierarchy>
<Hierarchy name="SubFamilia" hasAll="true" allMemberName="Subfamilias"
primaryKey="GenKey" primaryKeyTable="DimGen">
  <Table name="DimGen">
  </Table>
  <Level name="familia" table="DimGen" column="familia" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
  </Level>
  <Level name="subfamilia" table="DimGen" column="subfamilia"
type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
  </Level>
</Hierarchy>
</Dimension>
<Dimension type="StandardDimension" name="estructs">
  <Hierarchy name="molecularfunc" hasAll="true" allMemberName="estructs"
primaryKey="EstructuraKey" primaryKeyTable="DimEstructura">
  <Table name="DimEstructura">
  </Table>
  <Level name="molecularfunc" table="DimEstructura"
column="molecularfunc" type="String" uniqueMembers="false" levelType="Regular"
hideMemberIf="Never">
  </Level>
  <Level name="UI&#39;s" table="DimEstructura" column="uis" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
  </Level>
</Hierarchy>
<Hierarchy name="numui" hasAll="true" allMemberName="All ui"
primaryKey="EstructuraKey" primaryKeyTable="DimEstructura">
  <Table name="DimEstructura">
  </Table>
```

```

        <Level name="num_ui" table="DimEstructura" column="num_ui"
type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
        </Level>
        <Level name="Estructura" table="DimEstructura" column="EstructuraKey"
type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
        </Level>
        <Level name="funcion molecular" table="DimEstructura"
column="molecularfunc" type="String" uniqueMembers="true" levelType="Regular"
hideMemberIf="Never">
        </Level>
        <Level name="arreglo" table="DimEstructura" column="uis" type="String"
uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
        </Level>
    </Hierarchy>
</Dimension>
<Cube name="dwbiof" cache="true" enabled="true">
    <Table name="FactAnalysis_Fractal">
    </Table>
    <DimensionUsage source="DimChr" name="DimChr" foreignKey="ChrKey">
    </DimensionUsage>
    <Dimension type="StandardDimension" foreignKey="bandKey"
name="DimRange">
        <Hierarchy name="10partesX1unidad" hasAll="true"
allMemberName="intervalos_10_partes" primaryKey="bandkey">
            <Table name="DimDBand">
            </Table>
            <Level name="intervalosXunidad" table="DimDBand"
column="intervalo_1_unid" type="String" uniqueMembers="true" levelType="Regular"
hideMemberIf="Never">
            </Level>
            <Level name="10" table="DimDBand" column="10_limite_alto"
type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
            </Level>
        </Hierarchy>
        <Hierarchy name="20partesX1unidad" hasAll="true"
allMemberName="intervalos_20_partes" primaryKey="bandkey">
            <Table name="DimDBand">
            </Table>
            <Level name="intervalosXunidad" table="DimDBand"
column="intervalo_1_unid" type="String" uniqueMembers="true" levelType="Regular"
hideMemberIf="Never">
            </Level>
            <Level name="20" table="DimDBand" column="20_limite_alto"
type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
            </Level>
        </Hierarchy>
    </Dimension>

```

```

        <Hierarchy          name="30partesX1unidad"          hasAll="true"
allMemberName="intervalos_30_partes" primaryKey="bandkey">
        <Table name="DimDBand">
        </Table>
        <Level          name="intevalosXunidad"          table="DimDBand"
column="intervalo_1_unid"          type="String"          uniqueMembers="true"          levelType="Regular"
hideMemberIf="Never">
        </Level>
        <Level name="30" table="DimDBand" column="30_limite_alto"
type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
        </Level>
    </Hierarchy>
    <Hierarchy          name="40partesX1unidad"          hasAll="true"
allMemberName="intervalos_40_partes" primaryKey="bandkey">
        <Table name="DimDBand">
        </Table>
        <Level          name="intervalosXunidad"          table="DimDBand"
column="intervalo_1_unid"          type="String"          uniqueMembers="true"          levelType="Regular"
hideMemberIf="Never">
        </Level>
        <Level name="40" table="DimDBand" column="40_limite_alto"
type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
        </Level>
    </Hierarchy>
    <Hierarchy          name="50partesX1unidad"          hasAll="true"
allMemberName="intervalos_50_partes" primaryKey="bandkey">
        <Table name="DimDBand">
        </Table>
        <Level          name="intervalosXunidad"          table="DimDBand"
column="intervalo_1_unid"          type="String"          uniqueMembers="true"          levelType="Regular"
hideMemberIf="Never">
        </Level>
        <Level name="50" table="DimDBand" column="50_limite_alto"
type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
        </Level>
    </Hierarchy>
    <Hierarchy          name="60partesX1unidad"          hasAll="true"
allMemberName="intervalos_60_partes" primaryKey="bandkey">
        <Table name="DimDBand">
        </Table>
        <Level          name="intervalosXunidad"          table="DimDBand"
column="intervalo_1_unid"          type="String"          uniqueMembers="true"          levelType="Regular"
hideMemberIf="Never">
        </Level>
        <Level name="60" table="DimDBand" column="60_limite_alto"
type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
        </Level>
    </Hierarchy>
    
```

```
</Hierarchy>
  <Hierarchy          name="10partesX2unidad"          hasAll="true"
allMemberName="inetrvalos_en_10" primaryKey="bandkey">
  <Table name="DimDBand">
  </Table>
  <Level          name="intervalosX2unidades"          table="DimDBand"
column="intervalo_2_unid"          type="String"          uniqueMembers="true"          levelType="Regular"
hideMemberIf="Never">
  </Level>
  <Level name="10" table="DimDBand" column="10x2U_limite_alto"
type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
  </Level>
</Hierarchy>
  <Hierarchy          name="20partesX2unidad"          hasAll="true"
allMemberName="intervalos_en_20" primaryKey="bandkey">
  <Table name="DimDBand">
  </Table>
  <Level          name="20partesX2unidad"          table="DimDBand"
column="intervalo_2_unid"          type="String"          uniqueMembers="true"          levelType="Regular"
hideMemberIf="Never">
  </Level>
  <Level name="20" table="DimDBand" column="10_limite_alto"
type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
  </Level>
</Hierarchy>
  <Hierarchy          name="30partesX2unidad"          hasAll="true"
allMemberName="intervalos_30_partes" primaryKey="bandkey">
  <Table name="DimDBand">
  </Table>
  <Level          name="30partesX2unidades"          table="DimDBand"
column="intervalo_2_unid"          type="String"          uniqueMembers="true"          levelType="Regular"
hideMemberIf="Never">
  </Level>
  <Level name="30" table="DimDBand" column="30x2U_limite_alto"
type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
  </Level>
</Hierarchy>
  <Hierarchy          name="40partesX2unidad"          hasAll="true"
allMemberName="intervalos_40_partes" primaryKey="bandkey">
  <Table name="DimDBand">
  </Table>
  <Level          name="40partesX2unidades"          table="DimDBand"
column="intervalo_2_unid"          type="String"          uniqueMembers="true"          levelType="Regular"
hideMemberIf="Never">
  </Level>
  <Level name="40" table="DimDBand" column="20_limite_alto"
type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
```

```

        </Level>
    </Hierarchy>
    <Hierarchy          name="50partesX2unidad"          hasAll="true"
allMemberName="intervalos_50_partes" primaryKey="bandkey">
        <Table name="DimDBand">
        </Table>
        <Level          name="5partesX2unidades"          table="DimDBand"
column="intervalo_2_unid"          type="String"          uniqueMembers="true"          levelType="Regular"
hideMemberIf="Never">
        </Level>
        <Level name="50" table="DimDBand" column="50x2U_limite_alto"
type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
        </Level>
    </Hierarchy>
    <Hierarchy          name="60partesX2unidad"          hasAll="true"
allMemberName="intervalos_60_partes" primaryKey="bandkey">
        <Table name="DimDBand">
        </Table>
        <Level          name="60partesX2unidades"          table="DimDBand"
column="intervalo_2_unid"          type="String"          uniqueMembers="false"          levelType="Regular"
hideMemberIf="Never">
        </Level>
        <Level name="60" table="DimDBand" column="30_limite_alto"
type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
        </Level>
    </Hierarchy>
    <Hierarchy          name="10partesX3unidad"          hasAll="true"
allMemberName="intervalos_10_partes" primaryKey="bandkey">
        <Table name="DimDBand">
        </Table>
        <Level          name="10partesX10unidades"          table="DimDBand"
column="intervalo_10_unid"          type="String"          uniqueMembers="false"          levelType="Regular"
hideMemberIf="Never">
        </Level>
        <Level          name="10"          table="DimDBand"
column="10x10U_limite_alto"          type="String"          uniqueMembers="false"          levelType="Regular"
hideMemberIf="Never">
        </Level>
    </Hierarchy>
    <Hierarchy          name="20partesX3unidad"          hasAll="true"
allMemberName="intervalos_20_partes" primaryKey="bandkey">
        <Table name="DimDBand">
        </Table>
        <Level          name="20partesX10unidades"          table="DimDBand"
column="intervalo_10_unid"          type="String"          uniqueMembers="true"          levelType="Regular"
hideMemberIf="Never">
        </Level>
    </Hierarchy>
    </Level>

```

```

        <Level          name="20"          table="DimDBand"
column="20x10U_limite_alto"  type="String"  uniqueMembers="true"  levelType="Regular"
hideMemberIf="Never">
        </Level>
    </Hierarchy>
    <Hierarchy          name="30partesX3unidad"          hasAll="true"
allMemberName="intervalos_30_partes" primaryKey="bandkey">
        <Table name="DimDBand">
        </Table>
        <Level          name="30partesX10unidades"          table="DimDBand"
column="intervalo_10_unid"  type="String"  uniqueMembers="true"  levelType="Regular"
hideMemberIf="Never">
        </Level>
        <Level          name="30"          table="DimDBand"
column="30x10U_limite_alto"  type="String"  uniqueMembers="true"  levelType="Regular"
hideMemberIf="Never">
        </Level>
    </Hierarchy>
    <Hierarchy          name="40partesX3unidad"          hasAll="true"
allMemberName="intervalos_40_partes" primaryKey="bandkey">
        <Table name="DimDBand">
        </Table>
        <Level          name="40partesX10unidades"          table="DimDBand"
column="intervalo_10_unid"  type="String"  uniqueMembers="true"  levelType="Regular"
hideMemberIf="Never">
        </Level>
        <Level          name="40"          table="DimDBand"
column="40x10U_limite_alto"  type="String"  uniqueMembers="true"  levelType="Regular"
hideMemberIf="Never">
        </Level>
    </Hierarchy>
    <Hierarchy          name="50partesX3unidad"          hasAll="true"
allMemberName="intervalos_50_partes" primaryKey="bandkey">
        <Table name="DimDBand">
        </Table>
        <Level          name="50partesX10unidades"          table="DimDBand"
column="intervalo_10_unid"  type="String"  uniqueMembers="true"  levelType="Regular"
hideMemberIf="Never">
        </Level>
        <Level          name="50"          table="DimDBand"          column="10_limite_alto"
type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
        </Level>
    </Hierarchy>
    <Hierarchy          name="60partesX3unidad"          hasAll="true"
allMemberName="intervalos_60_partes" primaryKey="bandkey">
        <Table name="DimDBand">
        </Table>
    
```

```

        <Level      name="60partesX10unidades"      table="DimDBand"
column="intervalo_10_unid"  type="String"      uniqueMembers="true"      levelType="Regular"
hideMemberIf="Never">
        </Level>
        <Level      name="60"                      table="DimDBand"
column="60x10U_limite_alto"  type="String"      uniqueMembers="true"      levelType="Regular"
hideMemberIf="Never">
        </Level>
        </Hierarchy>
        </Dimension>
        <DimensionUsage source="DimGen" name="DimGen" foreignKey="GenKey">
        </DimensionUsage>
        <DimensionUsage source="estructs" name="estructs" foreignKey="EstructuraKey">
        </DimensionUsage>
        <Measure      name="%Size"          column="Longitud"      datatype="String"
formatString="0#.####%" aggregator="avg">
        </Measure>
        <Measure      name="%#UI"          column="Num_unid_info"  datatype="String"
formatString="0#.#####0" aggregator="avg">
        </Measure>
        <Measure      name="%Simetria"      column="Simetria"      datatype="String"
formatString="0#.0###" aggregator="avg">
        </Measure>
        <Measure name="count_D" column="D" datatype="String" formatString="0#.####"
aggregator="count">
        </Measure>
        <Measure name="%R2" column="R2" datatype="String" formatString="0#.##"
aggregator="avg">
        </Measure>
        <Measure name="%D" column="D" datatype="String" formatString="0#.####%"
aggregator="avg" visible="true">
        </Measure>
        </Cube>
        <Cube name="dwbio_ui" cache="true" enabled="true">
        <Table name="FactAnalysis_ui">
        </Table>
        <DimensionUsage source="DimChr" name="DimChr" foreignKey="ChrKey">
        </DimensionUsage>
        <DimensionUsage source="DimGen" name="DimGen" foreignKey="GenKey">
        </DimensionUsage>
        <DimensionUsage source="estructs" name="estructs" foreignKey="EstructuraKey">
        </DimensionUsage>
        <Dimension type="StandardDimension" foreignKey="TipoKey" name="DimTipo">
        <Hierarchy name="DimTipo" hasAll="true" allMemberName="uis"
primaryKey="TipoKey">
        <Table name="DimTipo">
        </Table>
    
```

```

        <Level name="tipoid" table="DimTipo" column="Tipoid"
type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
        </Level>
        <Level name="tipodescripcion" table="DimTipo"
column="Descripcion" type="String" uniqueMembers="false" levelType="Regular"
hideMemberIf="Never">
        </Level>
    </Hierarchy>
</Dimension>
<Measure name="Size" column="Longitud" datatype="Integer"
formatString="#.##0" aggregator="avg">
</Measure>
<Measure name="Orden_estrut" column="Orden_estrut" datatype="Integer"
formatString="Standard" aggregator="sum" caption="">
</Measure>
<Measure name="Orden_tipo" column="descrip_orden_tipo" datatype="String"
formatString="Standard" aggregator="max" formatter="" caption="">
</Measure>
<Measure name="Pos_final" column="Pos_final" datatype="Integer"
formatString="Standard" aggregator="sum">
</Measure>
<Measure name="Pos_inicial" column="Pos_inicial" datatype="Integer"
formatString="Standard" aggregator="sum">
</Measure>
</Cube>
<VirtualCube enabled="true" name="fui">
</VirtualCube>
</Schema>
    
```

1.3. Querypivotchr.xaction

Este Action sequence recibe una consulta mdx como cadena de caracteres y se encarga de mostrarla en Jpivot, es a esta consulta mdx la consulta que se parametriza para crear los filtros y cumplir con los requerimientos del Data mart Análisis Fractal.

```

<?xml version="1.0" encoding="UTF-8"?>
<action-sequence>
  <title>Análisis Fractal</title>
  <version>1</version>
  <logging-level>ERROR</logging-level>
  <documentation>
    <author>yo</author>
    <description>QueryPivot.xaction ejecuta una consulta mdx que recibe como un
    parametro</description>
    <help/>
    <result-type>report</result-type>
    <icon/>
  </documentation>
    
```



```
<inputs>
  <mdx type="string">
    <sources>
      <request>mdx</request>
    </sources>
    <default-value><![CDATA[]]></default-value>
  </mdx>
  <mode type="string">
    <sources>
      <request>mode</request>
    </sources>
    <default-value/>
  </mode>
</inputs>

<outputs>
  <model type="string"/>
  <connection type="string"/>
  <mdx type="string"/>
  <options type="list"/>
  <title type="string"/>
  <url type="string">
    <destinations>
      <response>redirect</response>
    </destinations>
  </url>
  <chartaxistickfontsize type="string"/><chartslicerfontsize type="string"/><chartlegendfontfamily
type="string"/><showgrid type="string"/><actionreference type="string"/><chartaxislabelfontsize
type="string"/><chartslicerlocation type="string"/><chartaxistickfontfamily
type="string"/><charheight type="string"/><chartslicerfontfamily type="string"/><chartvertaxislabel
type="string"/><chartaxislabelfontstyle type="string"/><chartdrillthroughenabled
type="string"/><chartbackgrounddb type="string"/><chartwidth type="string"/><role
type="string"/><chartaxistickfontstyle type="string"/><charttitlefontsize
type="string"/><chartbackgrounddg type="string"/><chartlegendfontstyle
type="string"/><chartshowslicer type="string"/><charttitlefontstyle
type="string"/><chartlegendlocation type="string"/><chartlocation type="string"/><chartbackgroundr
type="string"/><charttype type="string"/><chartaxislabelfontfamily
type="string"/><chartlegendfontsize type="string"/><chartshowlegend
type="string"/><chartaxisticklabelrotation type="string"/><charthorizaxislabel
type="string"/><charttitle type="string"/><chartsliceralignment type="string"/><chartslicerfontstyle
type="string"/><charttitlefontfamily type="string"/>
</outputs>

<actions>
  <action-definition>
    <component-name>PivotViewComponent</component-name>
    <action-type>aaa</action-type>
    <action-inputs>
      <mdx type="string"/>
      <mode type="string"/>
    </action-inputs>
    <action-outputs>
      <model type="string"/>
      <connection type="string"/>
    </action-outputs>
  </action-definition>
</actions>
```

```
<mdx type="string"/>
<options type="list"/>
<title type="string"/>
<url type="string"/>
<chartaxistickfontsize type="string"/><chartslicerfontsize type="string"/><chartlegendfontfamily
type="string"/><showgrid type="string"/><actionreference type="string"/><chartaxislabelfontsize
type="string"/><chartslicerlocation type="string"/><chartaxistickfontfamily
type="string"/><charheight type="string"/><chartslicerfontfamily type="string"/><chartvertaxislabel
type="string"/><chartaxislabelfontstyle type="string"/><chartdrillthroughenabled
type="string"/><chartbackgrounddb type="string"/><chartwidth type="string"/><role
type="string"/><chartaxistickfontstyle type="string"/><charttitlefontsize
type="string"/><chartbackgrounddg type="string"/><chartlegendfontstyle
type="string"/><chartshowslicer type="string"/><charttitlefontstyle
type="string"/><chartlegendlocation type="string"/><chartlocation type="string"/><chartbackgroundr
type="string"/><charttype type="string"/><chartaxislabelfontfamily
type="string"/><chartlegendfontsize type="string"/><chartshowlegend
type="string"/><chartaxisticklabelrotation type="string"/><charthorizaxislabel
type="string"/><charttitle type="string"/><chartsliceralignment type="string"/><chartslicerfontstyle
type="string"/><charttitlefontfamily type="string"/></action-outputs>
<component-name>PivotViewComponent</component-name>
<component-definition>
<title>Análisis Fractal</title>
<viewer>Pivot</viewer>
<model>solution:samples/analysis/DwBiofinal.mondrian.xml</model>
<jndi>dwbiofinal</jndi>
<query><![CDATA[{mdx}]]></query>
<options>
<personal/>
<cube-nav/>
<mdx-edit/>
<sort-conf/>
<level-style/>
<hide-spans/>
<properties/>
<non-empty/>
<swap-axes/>
<drill-member/>
<drill-position/>
<drill-replace/>
<drill-thru/>
<chart/>
<chart-conf/>
<print-conf/>
<print-pdf/>
<excel/>
</options>
<chartaxistickfontsize>10</chartaxistickfontsize><chartslicerfontsize>12</chartslicerfontsize><chart
legendfontfamily>SansSerif</chartlegendfontfamily><showgrid>true</showgrid><actionreference>s
amples/analysis/QueryPivotchr.xaction</actionreference><chartaxislabelfontsize>12</chartaxislabe
lfontsize><chartslicerlocation>1</chartslicerlocation><chartaxistickfontfamily>SansSerif</chartaxisti
ckfontfamily><charheight>300</charheight><chartslicerfontfamily>SansSerif</chartslicerfontfamily
><chartvertaxislabel></chartvertaxislabel><chartaxislabelfontstyle>0</chartaxislabelfontstyle><char
tdrillthroughenabled>false</chartdrillthroughenabled><chartbackgrounddb>255</chartbackgrounddb>
<chartwidth>1200</chartwidth><role/><chartaxistickfontstyle>0</chartaxistickfontstyle><charttitlefo
```

```
ntsize>18</charttitlefontsize><chartbackgroundg>255</chartbackgroundg><mdx>with member
[Measures].[#D's X familia] as 'Count(Filter([DimGen.Familia].[Familias].Children,
([Measures].[%R2] &gt; 0.9)))'
  member [Measures].[#D's X gen] as 'Count(Filter([DimGen.genid].[genes].Children,
([Measures].[%R2] &gt; 0.9)))'
  member [Measures].[GenID] as '[DimGen].CurrentMember.Name'
select NON EMPTY {[Measures].[%#UI], [Measures].[%D], [Measures].[#D's X familia],
[Measures].[#D's X gen]} ON COLUMNS,
  Filter(Hierarchize(Union({[DimRange.DimDBand.range40].[rangos de 40],
[DimRange.DimDBand.range40].[rangos de 40].[2 [1 - 1.5]]}),
[DimRange.DimDBand.range40].[rangos de 40].[2 [1 - 1.5]].Children)), ([Measures].[%#UI] &gt;
4.0)) ON ROWS
from [dwbiof]
</mdx><chartlegendfontstyle>0</chartlegendfontstyle><chartshowslicer>true</chartshowslicer><ch
arttitlefontstyle>1</charttitlefontstyle><connection/><chartlegendlocation>3</chartlegendlocation><
chartlocation>none</chartlocation><chartbackgroundr>255</chartbackgroundr><charttype>1</char
ttype><chartaxislabelfontfamily>SansSerif</chartaxislabelfontfamily><chartlegendfontsize>10</cha
rtlegendfontsize><chartshowlegend>true</chartshowlegend><chartaxisticklabelrotation>90</charta
xisticklabelrotation><charthorizontalaxislabel></charthorizontalaxislabel><charttitle></charttitle><chartsliceral
ignment>3</chartsliceralignment><chartslicerfontstyle>0</chartslicerfontstyle><charttitlefontfamily>
SansSerif</charttitlefontfamily></component-definition>
  <action-name>Pivot View</action-name>
  <logging-level>DEBUG</logging-level>
</action-definition>

</actions>
</action-sequence>
```

1.4. Consultas mdx

Aquí se muestran las consultas mdx usadas para la solución de los dos mdx.

1.4.1. Análisis FractalAnalysis

```
WITH member [Measures].[#D's X familia] as 'Count(Filter([DimGen.Familia],[Familias].Children, ([Measures].[%R2] > 0.9)))'
```

```
member [Measures].[#D's X gen] as 'Count(Filter([DimGen.genid],[genes].Children, ([Measures].[%R2] > 0.9)))'
```

```
member [Measures].[GenID] as '[DimGen].CurrentMember.Name'
```

```
SELECT NON EMPTY {[Measures].[%D], [Measures].[#D's X familia], [Measures].[#D's X gen], [Measures].[%#UI], [Measures].[%Size], [Measures].[%Simetria], [Measures].[count_D], [Measures].[%R2]} ON COLUMNS,
```

```
NON EMPTY [DimRange.40partesX1unidad].[intervalos_40_partes].[1 [0 - 1]].Children ON ROWS
```

```
FROM [dwbiof]
```

```
WHERE [DimChr.Dimchr].[AllOrganisms].[homo_sapiens]
```

1.4.2. Unidades de información

```
SELECT NON EMPTY {[Measures].[Size], [Measures].[Orden_tipo], [Measures].[Pos_final], [Measures].[Pos_inicial]} ON COLUMNS,
```

```
NON EMPTY Order({[DimGen.genid].[genes], [estructs.numui].[All ui], [DimTipo.DimTipo].[uis]}, [Measures].[Size], DESC) ON ROWS
```

```
from [dwbio_ui]
```

```
where [DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr21]
```

1.4.3. Grafica de cromosomas vs medidas

```
with member [Measures].[#D's X familia] as 'Count(Filter([DimGen.Familia],[Familias].Children, ([Measures].[%R2] > 0.9)))'
```

```
member [Measures].[#D's X gen] as 'Count(Filter([DimGen.genid],[genes].Children, ([Measures].[%R2] > 0.9)))'
```

```
member [Measures].[GenID] as '[DimGen].CurrentMember.Name'
```

```
member [Measures].[long] as '[Measures].[%Size] - 1 - (-1)'
```

member [Measures].[ds] as '[Measures].[%D] - 1 - (-1)'

member [Measures].[sim] as '[Measures].[%Simetria] - 1 - (-1)'

member [Measures].[num_unid] as '[Measures].[%#UI] - 1 - (-1)'

member [Measures].[cuantos_d] as '[Measures].[Count_d] - 1 - (-1)'

```
select NON EMPTY Crossjoin({[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr01],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr02],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr03],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr04],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr05],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr06],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr07],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr08],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr09],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr10],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr11],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr12],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr13],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr14],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr15],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr16],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr17],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr18],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr19],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr20],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr21],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[Chr22],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[ChrMT],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[ChrX],  
[DimChr.Dimchr].[AllOrganisms].[homo_sapiens].[ChrY]}, {[Measures].[#D's X familia]}) ON  
COLUMNS,  
  
NON EMPTY {[DimRange.40partesX2unidad].[intervalos_40_partes].[1 [0 - 2]]}.children} ON  
ROWS  
  
from [dwbiof]
```

ANEXO J: SCRIPT LA BODEGA DE DATOS

Script de creación de la DW según el modelo dimensional.

```
CREATE DATABASE IF NOT EXISTS DwBiofinal;

USE DwBiofinal;

/*DIMENSION*/

/* Create table DimChr */

DROP TABLE IF EXISTS DimChr;

CREATE TABLE DimChr (
    ChrKey INT NOT NULL AUTO_INCREMENT PRIMARY KEY
, ChrID varchar(5) NULL
, Division varchar(6) NULL
, Organismo varchar(200) NULL
);

/*DIMENSION*/

/* Create table DimGen */

DROP TABLE IF EXISTS DimGen;

CREATE TABLE DimGen (
    GenKey INT NOT NULL AUTO_INCREMENT PRIMARY KEY
, GenID varchar(20) NULL
, descrip varchar(300) NULL
, familia varchar(50) NULL
, subfamilia varchar(50) NULL
, banda varchar(1) NULL
, inicio int NULL
, final int NULL
);

/*DIMENSION*/
```

```
/* Create table DimEstructura */  
  
DROP TABLE IF EXISTS DimEstructura;  
  
CREATE TABLE DimEstructura (  
    EstructuraKey int NOT NULL AUTO_INCREMENT PRIMARY KEY  
    , EstructuraID varchar(20) NULL  
    , uis varchar(4266) NULL  
    , num_ui int NULL  
    , molecularfunc varchar(1500) NULL  
);  
  
/*T A B L A - H E C H O S*/  
  
/* Create table FactAnalisis_Fractal*/  
  
DROP TABLE IF EXISTS FactAnalisis_Fractal;  
  
CREATE TABLE FactAnalisis_Fractal (  
    GenKey int NOT NULL  
    , EstructuraKey int NOT NULL  
    , ChrKey int NOT NULL  
    , bandKey int NOT NULL  
    , D NUMERIC(24,12) NULL  
    , R2 NUMERIC(24,12) NULL  
    , Num_unid_info int NULL  
    , Simetria NUMERIC(25, 11) NULL  
    , Longitud NUMERIC(7) NULL  
);  
  
/*D I M E N S I O N*/  
  
/* Create table DimTipo */  
  
DROP TABLE IF EXISTS DimTipo;  
  
CREATE TABLE DimTipo (  

```

```
TipoKey INT NOT NULL AUTO_INCREMENT PRIMARY KEY
, TipoID varchar(10) NOT NULL
, Descripcion varchar(20) NOT NULL
);
```

```
/*T A B L A - H E C H O S*/
```

```
/* Create table FactAnalisis_de_Unidades_de_Informacion*/
```

```
DROP TABLE IF EXISTS FactAnalisis_ui;
```

```
CREATE TABLE FactAnalisis_ui (
```

```
    TipoKey int NOT NULL
, GenKey int NOT NULL
, EstructuraKey int NOT NULL
, ChrKey int NOT NULL
, descrip_orden_tipo varchar(10)
, Orden_estrut int NULL
, Orden_tipo int NULL
, Pos_inicial int NULL
, Pos_final int NULL
, Longitud int NULL
);
```

```
/*Create table range*/
```

```
DROP TABLE IF EXISTS DimDBand;
```

```
CREATE TABLE DimDBand (
```

```
    bandkey int NOT NULL AUTO_INCREMENT PRIMARY KEY
, intervalo_1_unid varchar(30) NOT NULL
, intervalo_2_unid varchar(30) NOT NULL
, intervalo_10_unid varchar(30) NOT NULL
```


, band_lower_value float NOT NULL
, band_upper_value float NOT NULL
, band_div10 int NOT NULL
, limits_div10 varchar(39) NOT NULL
, band_div20 int NOT NULL
, band_div30 int NOT NULL
, band_div40 int NOT NULL
, band_div50 int NOT NULL
, band_div60 int NOT NULL
, limits_div10X2U varchar(39)NOT NULL
, limits_div20 varchar(39) NOT NULL
, limits_div30 varchar(39) NOT NULL
, limits_div40 varchar(39) NOT NULL
, limits_div50 varchar(39)NOT NULL
, limits_div60 varchar(39) NOT NULL
, 10_limite_bajo varchar(39) NOT NULL
, 10_limite_alto varchar(39) NOT NULL
, 20_limite_bajo varchar(39) NOT NULL
, 20_limite_alto varchar(39) NOT NULL
, 30_limite_bajo varchar(39) NOT NULL
, 30_limite_alto varchar(39) NOT NULL
, 40_limite_bajo varchar(39) NOT NULL
, 40_limite_alto varchar(39) NOT NULL
, 50_limite_bajo varchar(39) NOT NULL
, 50_limite_alto varchar(39) NOT NULL
, 60_limite_bajo varchar(39) NOT NULL
, 60_limite_alto varchar(39) NOT NULL

, low_string varchar(39) NOT NULL
, high_string varchar(39) NOT NULL
, 10x2U_limite_alto varchar(39) NOT NULL
, limits_div20_2_unids varchar(39) NOT NULL
, 20_parts_2_unids varchar(3) NOT NULL
, limits_div30_2_unids varchar(39) NOT NULL
, 30x2U_limite_alto varchar(39) NOT NULL
, limits_div40_2_unids varchar(39) NOT NULL
, 40_parts_2_unids varchar(3) NOT NULL
, limits_div50_2_unids varchar(39) NOT NULL
, 50_parts_2_unids varchar(3) NOT NULL
, 50x2U_limite_alto varchar(39) NOT NULL
, limits_div60_2_unids varchar(39) NOT NULL
, 60_parts_2_unids varchar(3) NOT NULL
, limits_div10_10_unids varchar(39) NOT NULL
, 10_parts_10_unids varchar(3) NOT NULL
, 10x10U_limite_alto varchar(39) NOT NULL
, limits_div20_10_unids varchar(39) NOT NULL
, 20_parts_10_unids varchar(3) NOT NULL
, 20x10U_limite_alto varchar(39) NOT NULL
, limits_div30_10_unids varchar(39) NOT NULL
, 30_parts_10_unids varchar(3) NOT NULL
, 30x10U_limite_alto varchar(39) NOT NULL
, limits_div40_10_unids varchar(39) NOT NULL
, 40_parts_10_unids varchar(3) NOT NULL
, 40x10U_limite_alto varchar(39) NOT NULL
, limits_div50_10_unids varchar(39) NOT NULL

```
, 50_parts_10_unids varchar(3)NOT NULL
, 50x10U_limite_alto varchar(39) NOT NULL
, limits_div60_10_unids varchar(39) NOT NULL
, 60_parts_10_unids varchar(3)NOT NULL
, 60x10U_limite_alto varchar(39) NOT NULL
);
```

```
/*Keys UI*//*Tipokey*/
```

```
ALTER TABLE FactAnalisis_ui ADD CONSTRAINT
FK_FactAnalisis_ui_TipoKey FOREIGN KEY
(
TipoKey
) REFERENCES DimTipo
(TipoKey )
ON UPDATE NO ACTION
ON DELETE NO ACTION;
```

```
/*Genkey*/
```

```
ALTER TABLE FactAnalisis_ui ADD CONSTRAINT
FK_FactAnalisis_ui_GenKey FOREIGN KEY
(
GenKey
) REFERENCES DimGen
(GenKey )
ON UPDATE NO ACTION
ON DELETE NO ACTION;
```

```
/*estructurakey*/
```

```
ALTER TABLE FactAnalysis_ui ADD CONSTRAINT  
FK_FactAnalysis_ui_EstructuraKey FOREIGN KEY  
(  
EstructuraKey  
) REFERENCES DimEstructura  
( EstructuraKey )  
ON UPDATE NO ACTION  
ON DELETE NO ACTION;
```

*/*chrkey*/*

```
ALTER TABLE FactAnalysis_ui ADD CONSTRAINT  
FK_FactAnalysis_ui_ChrKey FOREIGN KEY  
(  
ChrKey  
) REFERENCES DimChr  
( ChrKey )  
ON UPDATE NO ACTION  
ON DELETE NO ACTION;
```

/ Keys fractal*//*Genkey*/*

```
ALTER TABLE FactAnalysis_Fractal ADD CONSTRAINT  
FK_FactAnalysis_Fractal_GenKey FOREIGN KEY  
(  
GenKey  
) REFERENCES DimGen  
( GenKey )  
ON UPDATE NO ACTION
```

```
ON DELETE NO ACTION;

/*Estructurakey*/

ALTER TABLE FactAnalysis_Fractal ADD CONSTRAINT

FK_FactAnalysis_Fractal_EstructuraKey FOREIGN KEY

(

EstructuraKey

) REFERENCES DimEstructura

(EstructuraKey )

ON UPDATE NO ACTION

ON DELETE NO ACTION

;

/*Chrkey*/

ALTER TABLE FactAnalysis_Fractal ADD CONSTRAINT

FK_FactAnalysis_Fractal_ChrKey FOREIGN KEY

(

ChrKey

) REFERENCES DimChr

(ChrKey )

ON UPDATE NO ACTION

ON DELETE NO ACTION;

ALTER TABLE FactAnalysis_Fractal ADD CONSTRAINT

FK_FactAnalysis_Fractal_bandKey FOREIGN KEY

(

bandkey

) REFERENCES DimDBand

(bandkey )

ON UPDATE NO ACTION
```

ON DELETE NO ACTION;

ANEXO K: COMPARACION DE RESULTADO DE CONSULTAS CON DATOS CREADOS POR LOS INVESTIGADORES.

Este anexo contiene la comparación realizada entre los datos que se obtuvieron de la DW y los datos que obtuvieron los investigadores del grupo BIMAC con el método de scripts.

Para poder desarrollar la aplicación de usuario final, es necesario entender que los requerimientos de los investigadores sobrepasaban los límites de funcionalidad de las herramientas OLAP, por tal razón se hizo necesario construir un conjunto de filtros los cuales permiten a los usuarios poder seleccionar el organismo, todos o un cromosoma de dicho organismo, seleccionar un valor de R2 que permita visualizar todas las medidas que tengan un R2 superior al determinado, posteriormente el usuario podrá determinar un intervalo, el numero de partes en que se dividirá dicho intervalo, finalmente el usuario podrá graficar los resultados con la posibilidad de escoger que columna desea y en que eje.

Para el desarrollo de la aplicación de usuario final se hizo uso de Jpivot, herramienta perteneciente a la suite de Pentaho que permite mostrar los datos resultantes de las consultas, las cuales son procesadas por el servidor de consultas mdx Mondrian. A continuación se ilustrara el proceso de desarrollo de la aplicación, las actividades que se realizaron son las siguientes.

1. Teniendo en cuenta los requerimientos del cliente, fue necesario construir un servlet que permitiera crear el conjunto de filtros que capturaran los valores seleccionados por los usuarios, construyera una consulta mdx apartir de dichos valores y enviarla a Jpivot para que se encargase de mostrar los datos resultantes, supliendo así las carencias de funcionalidad de Jpivot.

Cromosoma-(Homo Sapiens)	
ChrAll ▾	
seleccione un R2	
0.90 ▾	
Seleccione el Intervalo	
<input checked="" type="radio"/> 1 unidad <input type="radio"/> 2 unidades <input type="radio"/> 10 unidades	
0 - 1 ▾	
Numero de partes	
40 ▾	
Ancho	Altura
900	500
EJE X	EJE Y
Seleccione... ▾	Seleccione... ▾

Figura k1. Conjunto de Filtros

Se necesitaba que los filtros y la presentación de los datos estuvieran en la misma ventana, además permitirle al usuario seleccionar que columna o columnas deseaba graficar y en que eje.

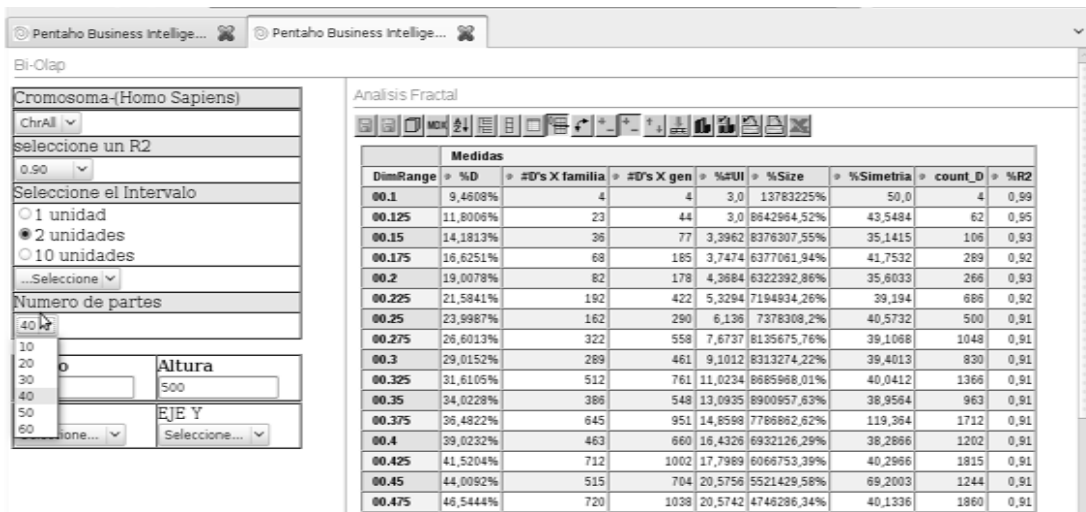


Figura K2. Analisis Fractal

Test de comparación entre los datos cargados en la DW y los datos que manejaban los investigadores del grupo BIMAC graficados en Excel después de extraerlos de los múltiples archivos con scripts.

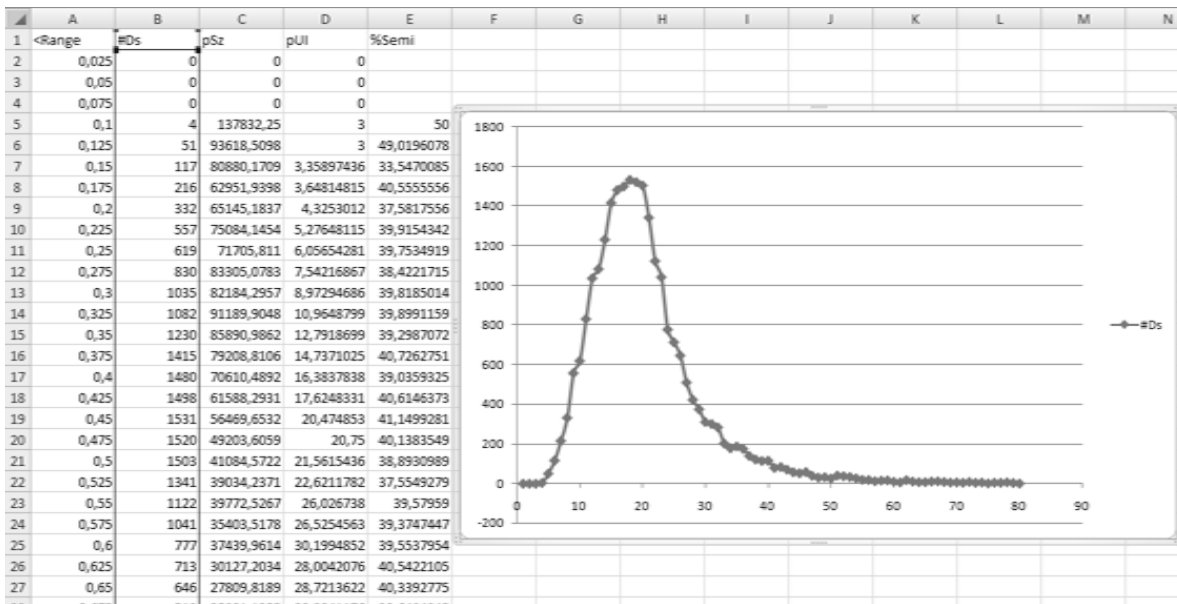


Figura K3. Rango vs Unidades de información en Excel.

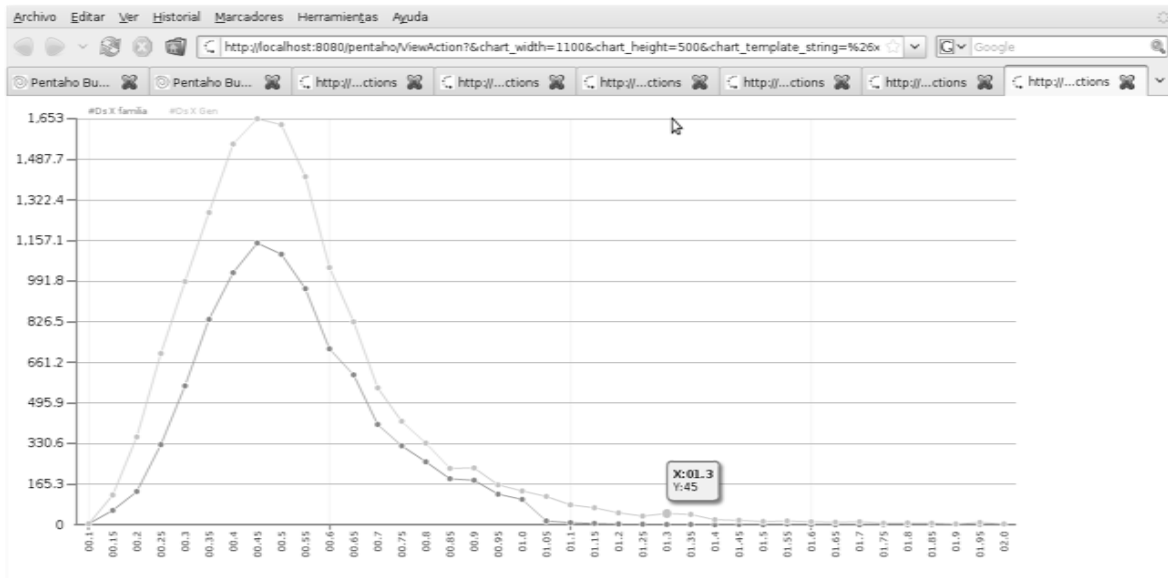


Figura K4. Rango vs número de unidades de información por gen y familia de genes en la DW.

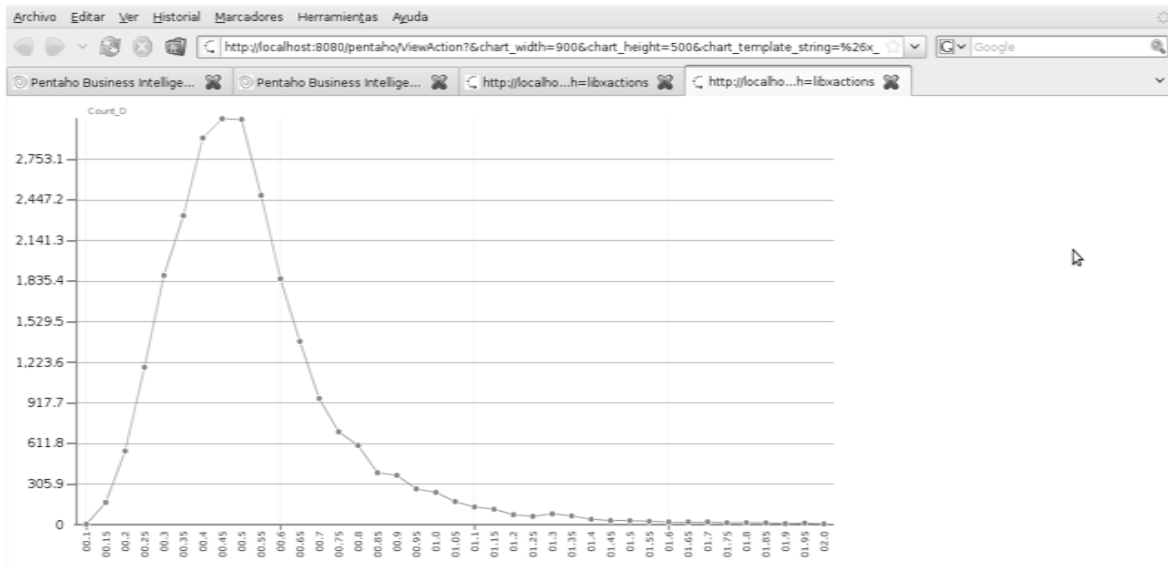


Figura K5. Rango vs Numero de unidades de información total en la DW.

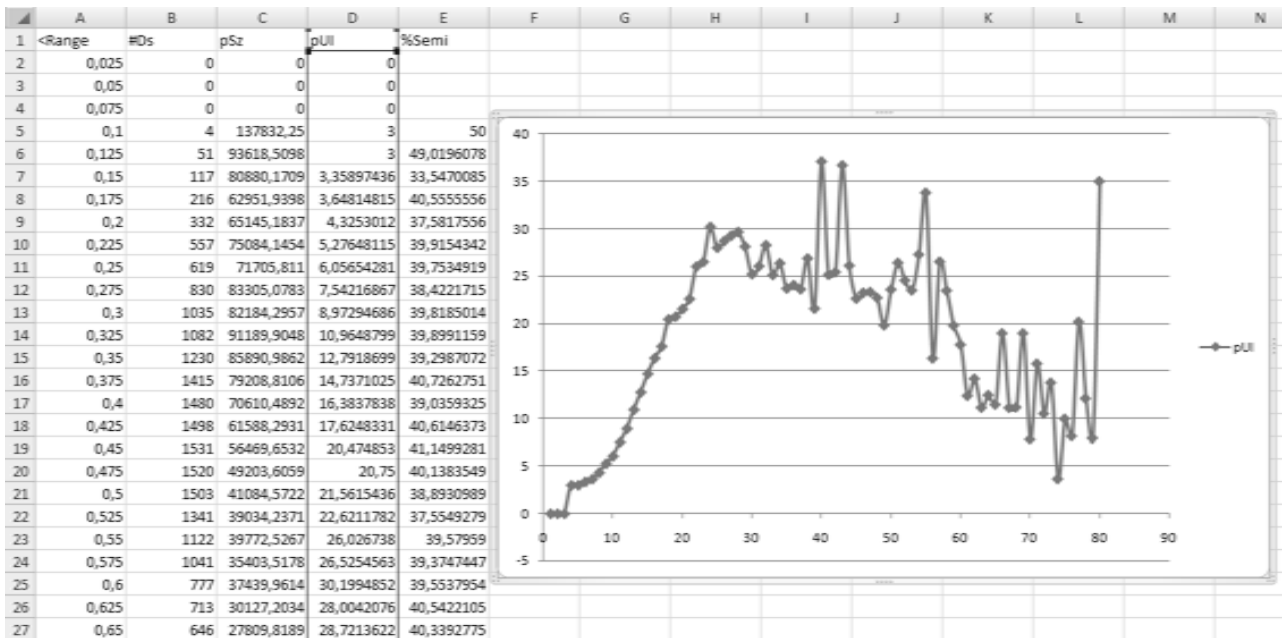


Figura K6. Rangos vs número Unidades de información en Excel.

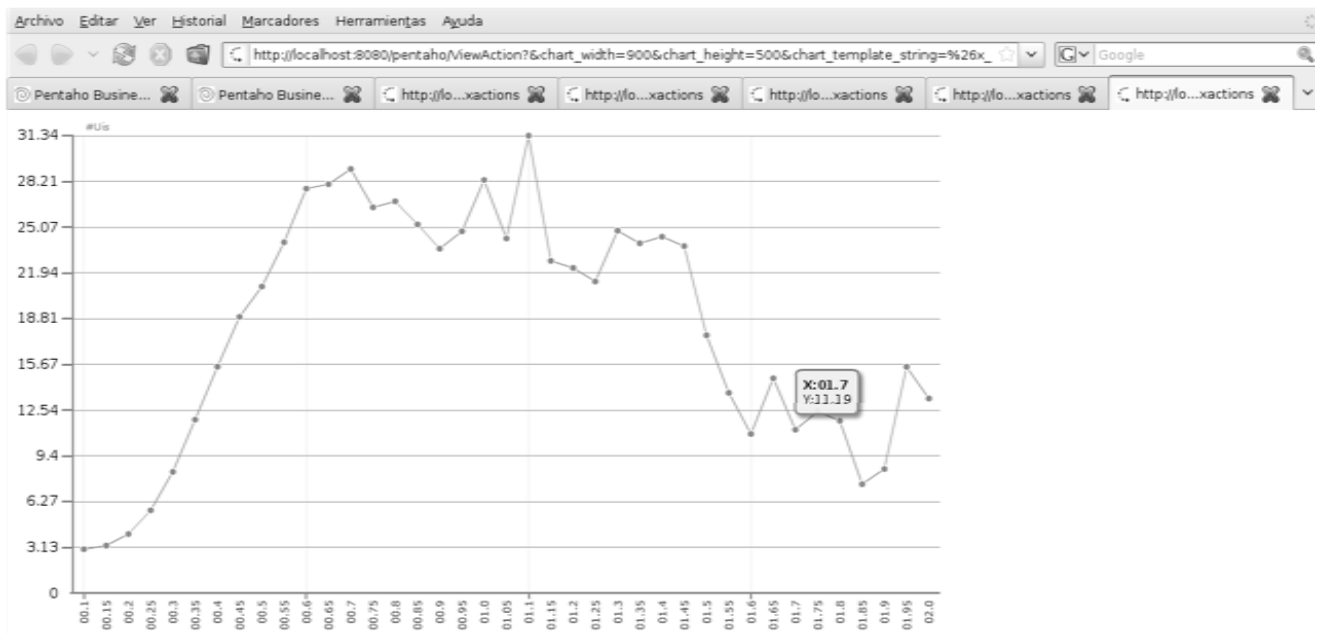


Figura K7. Rango vs número de unidades de información en la DW

El segundo Data Mart no requería la implementación de filtros especializados para su desarrollo aunque la presentación final de los datos que los investigadores deseaban hizo necesario tomar algunas consideraciones en los datos al momento de cargarlos puesto que no es habitual en herramientas OLAP que muestren cadenas de caracteres en las columnas de medida, generalmente son números sobre los que se realiza una operación de agregación ya sea suma, promedio, conteo, máximo o mínimo, sin embargo fue posible satisfacer los requerimientos de los usuarios haciendo uso de Jpivot y una consulta mdx.

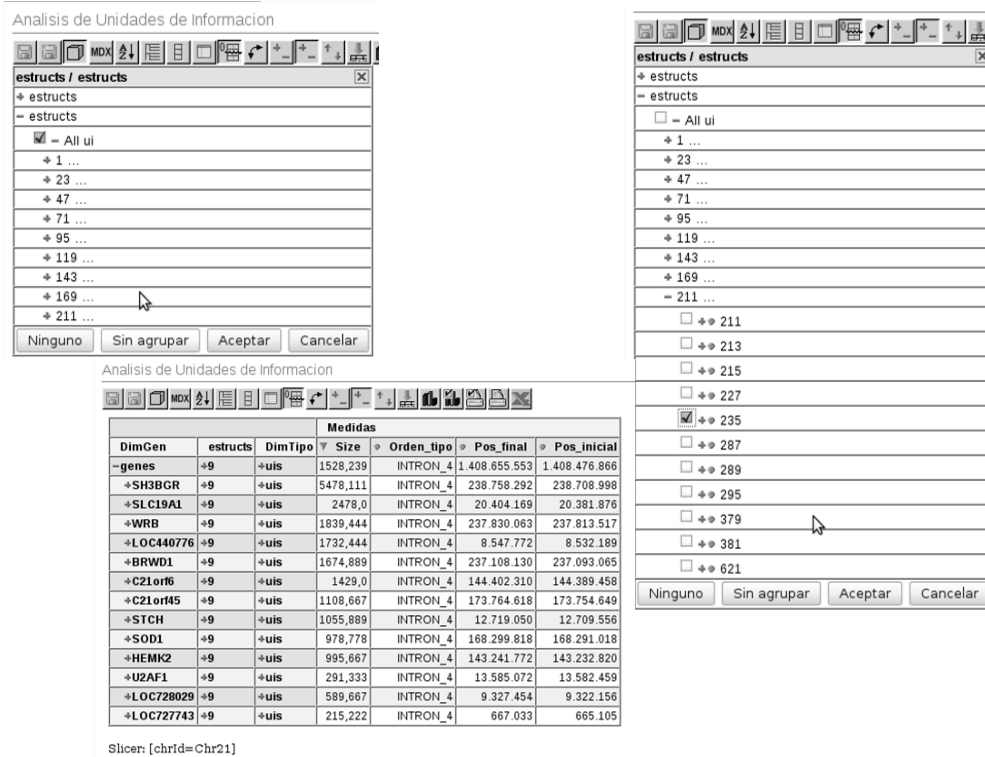


Figura K8. Análisis unidades de información

Adicionalmente se construyo una tercera consulta derivada del primer Data Mart en la cual los investigadores pudiesen comparar gráficamente las medidas en los intervalos con la precisión deseada, en todos los cromosomas y poder seleccionar cuales deseaban visualizar y cuáles no, brindandole los Filtros del primer Data Mart.

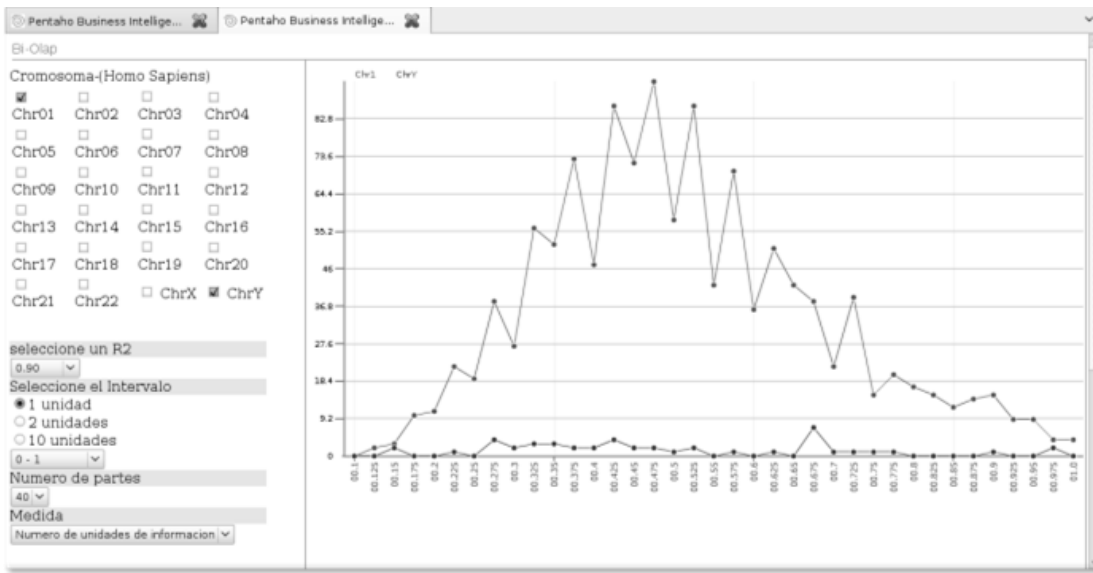


Figura K9.Cromosomas vs Medidas 1

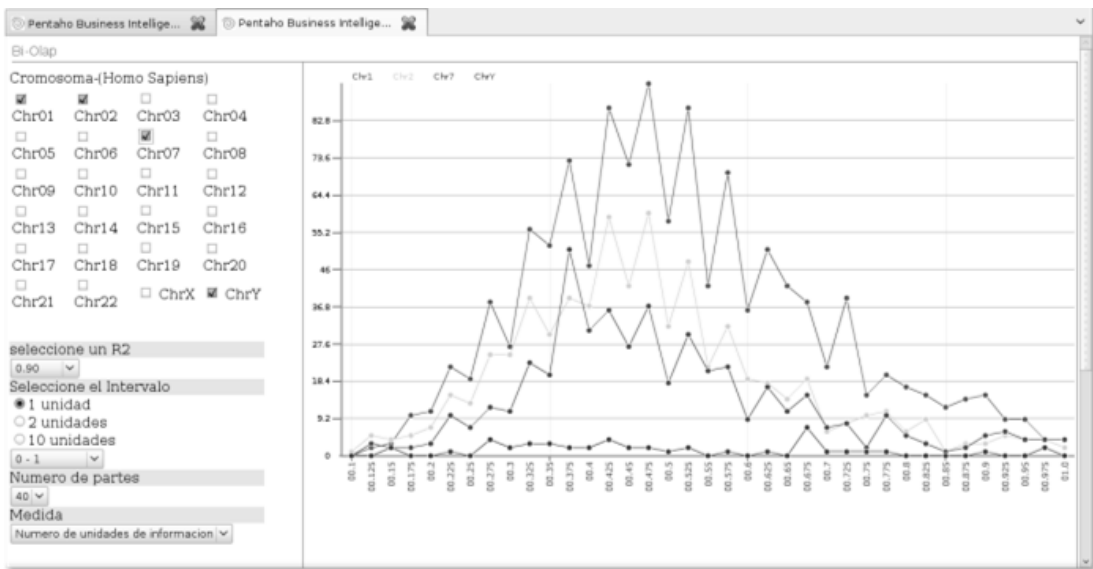


Figura K10. Cromosoma vs Medidas 2

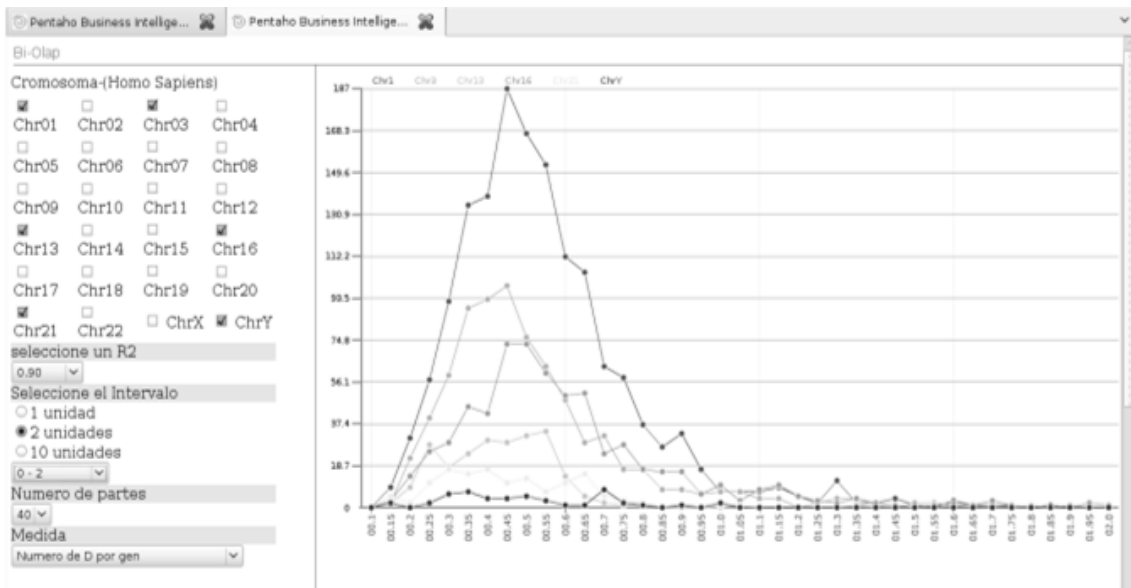


Figura K11. Cromosomas vs Medidas 3

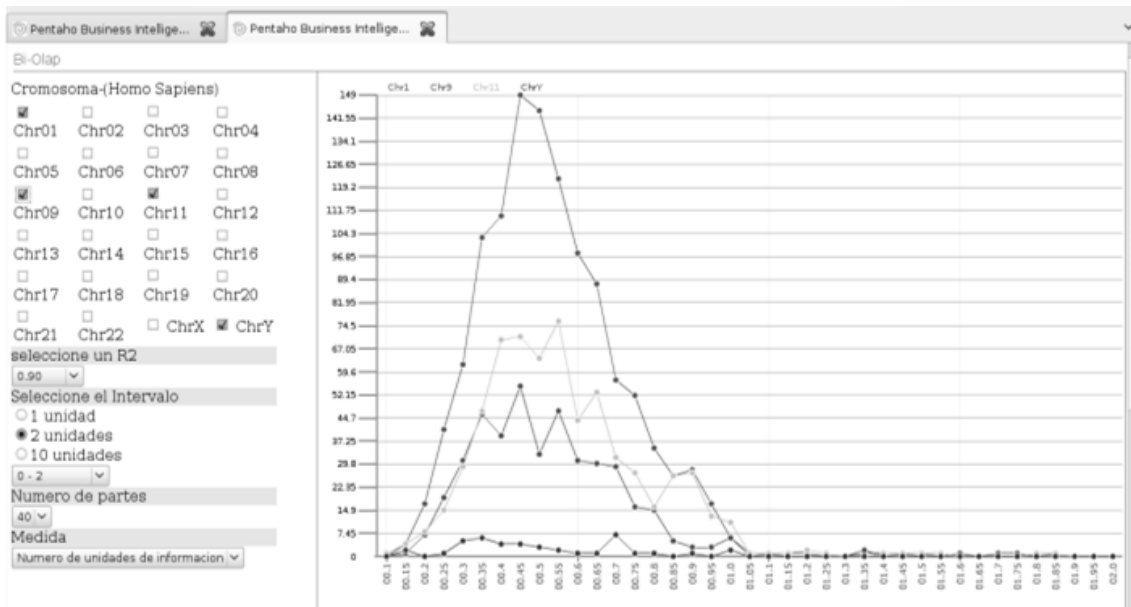


Figura K12. Cromosomas vs Medidas 4

ANEXO L: ARQUITECTURA DE HERRAMIENTAS USADAS

Es anexo muestra las arquitecturas de las dos herramientas más importantes que se usaron en la creación de la DW.

N.1. Arquitectura de Pentaho

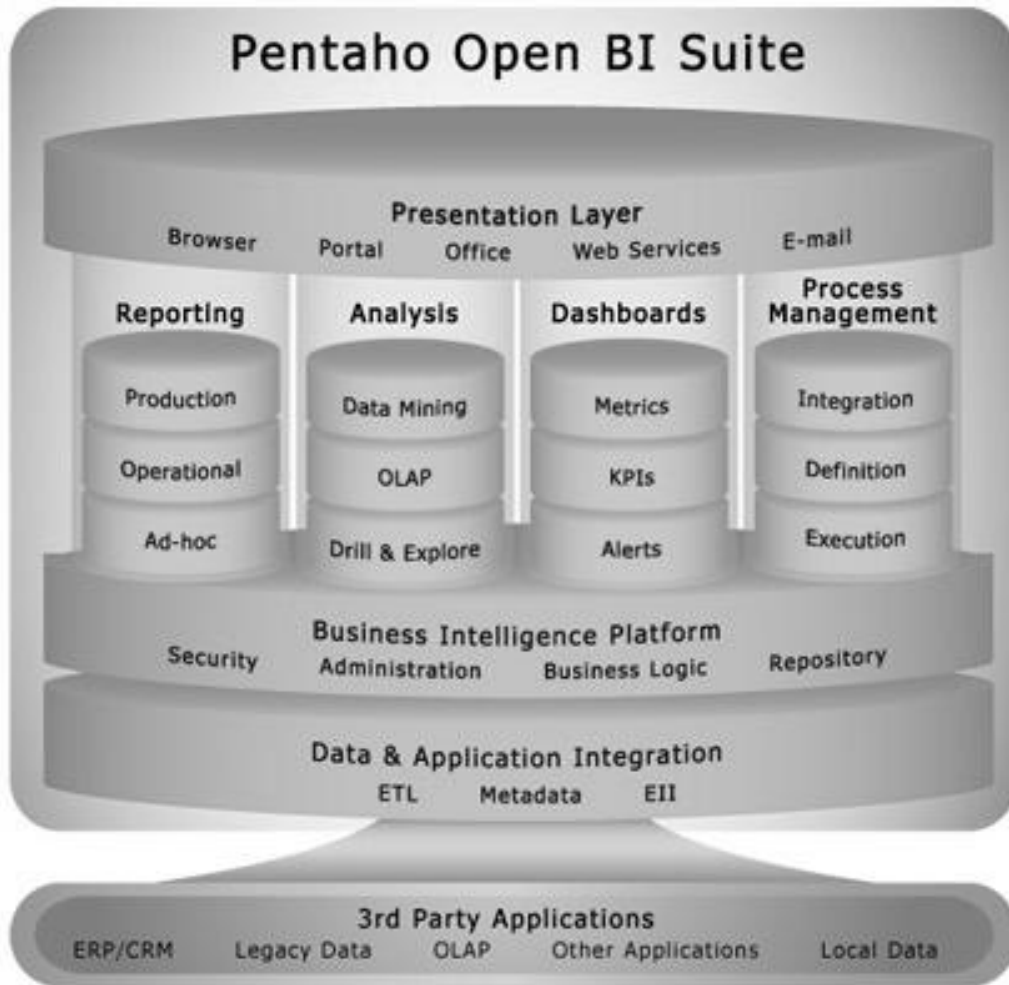


Figura L1. Arquitectura de Penatho

N.2. Arquitectura de Mondrian

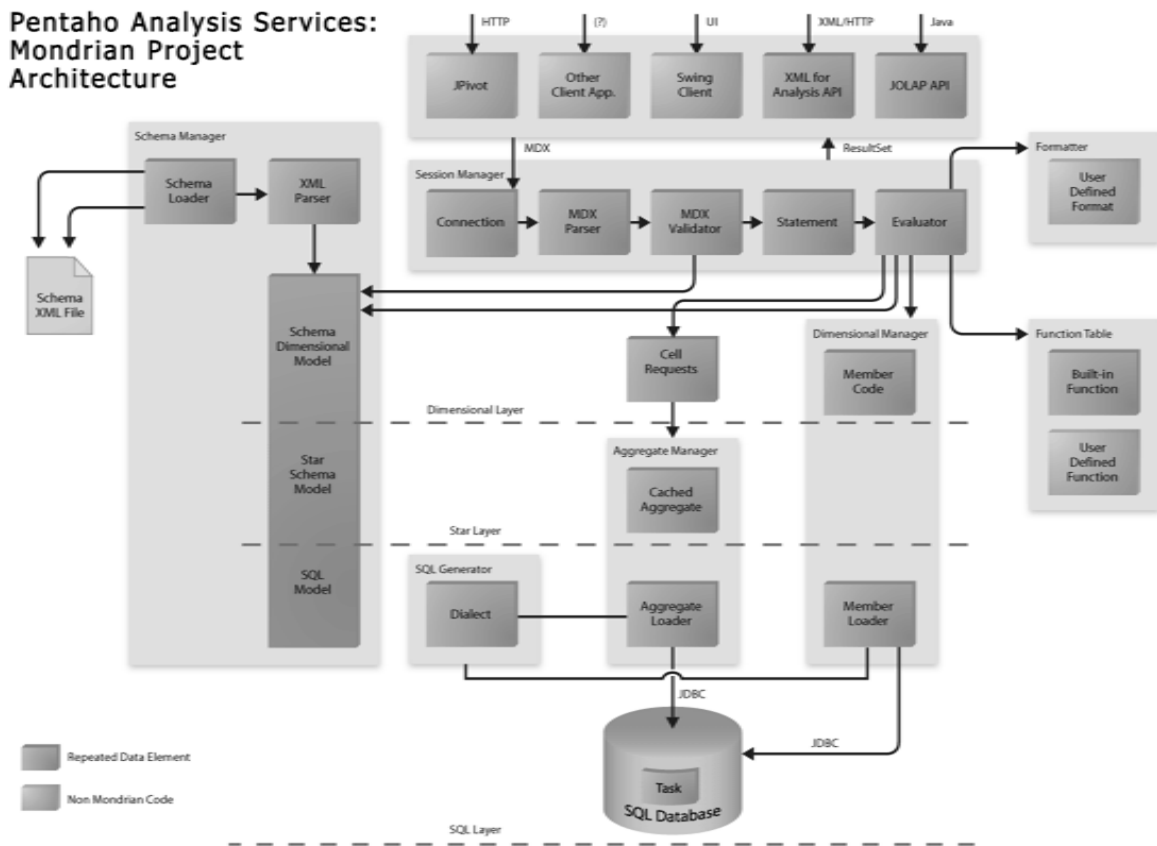


Figura L2. Arquitectura de Modrian

ANEXO M: EJEMPLO DE ARCHIVO GBK

Este anexo presenta un archivo de ejemplo de los que son descargados del Genbank, sobre los cuales el grupo BIMAC lleva a cabo sus investigaciones.

The image shows a screenshot of the NCBI GenBank website. The main content area displays the record for 'psbA [Ranunculus japonicus]'. The record includes the following information:

- LOCUS:** BAF43597 30 aa linear PLN 14-NOV-2008
- DEFINITION:** psbA [Ranunculus japonicus].
- ACCESSION:** BAF43597
- VERSION:** BAF43597.1 GI:121308427
- DBSOURCE:** accession AB244028.1
- KEYWORDS:** .
- SOURCE:** chloroplast Ranunculus japonicus
- ORGANISM:** Ranunculus japonicus
- REFERENCE 1:** Wang, R.-J., Cheng, C.-L., Chang, C.-C., Wu, C.-L., Su, T.-M. and Chaw, S.-M. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. BMC Evol. Biol. 8, 36 (2008). PMID: 18237435
- REFERENCE 2:** Wang, R.-J., Cheng, C.-L. and Chaw, S.-M. Direct Submission. Submitted (08-DEC-2005) Contact: Shu-Hsiaw Chaw Biodiversity center, Academia Sinica: Academia Rd, Taipei 115, ROC
- FEATURES:**
 - source:** 1..30 /organism="Ranunculus japonicus" /organelle="plastid:chloroplast" /db_xref="taxon:286919" /country="Taiwan" /collected_by="Chiaolei Cheng" /identified_by="Chiaolei Cheng" /PCR_primers="fwd_name: rpi2-psbA-F3, fwd_seq: gytacacgycctagtagtagagg, rev_name: rpi2-psbA-R2, rev_seq: ctggtatggagctatgcaygarctaa"
 - Protein:** <1..30 /name="psbA"
 - Region:** <2..30 /region_name="Photo_RC" /note="Photosynthetic reaction centre protein: c108220" /db_xref="CDD:126361"
 - CDS:** 1..30 /gene="psbA" /coded_by="complement(AB244028.1:612..>704)" /note="direction:right" /transl_table=11
- ORIGIN:** 1 cslgsevshe rnahnfpidi avevpsing

The right sidebar contains several sections:

- Change Region Shown:** A dropdown menu.
- Customize View:** A dropdown menu.
- Sequence Analysis Tools:** Includes BLAST Sequence (Find regions of similarity...), Conserved Domains (View conserved domains...), and Recent Activity (Turn Off, Clear).
- All links from this record:** A list of links including BLINK, Conserved Domains, Full text in PMC, Nucleotide, Related Structure, Related Sequences, Domain Relatives, PubMed, and Taxonomy.

At the bottom of the page, there is a footer with the following text:

Write to the Help Desk
NCBI | NLM | NIH
Department of Health & Human Services
Privacy Statement | Freedom of Information Act | Disclaimer

Figura M1. Archivo con formato gbk de ejemplo del Genbank

ANEXO N: PROBLEMAS DE CARGUE DE DATOS BIOLÓGICOS

El siguiente anexo presenta una clasificación de artefactos de datos biológicos

Una clasificación de artefactos de datos biológicos

1,2Judice L.Y. Koh, 2Mong Li Lee, 1Vladimir Brusic

1Knowledge Discovery Department, Institute for Infocomm Research

2School of Computing, National University of Singapore

Workshop on Database Issues in Biological Databases (DBiBD)

January 8-9, 2005

[National e-Science Centre](#), Edinburgh, Scotland, UK

in conjunction with [ICDT](#)

Jointly organized by the [European Bioinformatics Institute](#) and the [Edinburgh Database Group](#).

El análisis y publicación de bases de datos especializadas (SSDW) tarda 2 semanas a 1 mes. Pero por lo general, gastan hasta $\frac{1}{4}$ de su tiempo en la limpieza de los registros de datos.

Las bases de datos públicas (GENBANK, SWISS-PROT, DDBJ, EMBL, PIR, entre otras) proveen un gran recurso de fuentes de datos biológicos.

La precisión del análisis de datos y la capacidad para producir resultados correctos de minería de datos se basa en la calidad de los datos.

Pero como asegurar una alta calidad en los datos dentro de nuestra DW?

N.3. Objetivos de nuestro estudio:

Artefactos de datos biológicos en las bases de datos biológicos

Evaluación crítica de la calidad de los datos que los biólogos / científicos han estado utilizando para el análisis de datos y minería de datos.

Plan de trabajo a la mejora de la calidad de los datos en bases de datos moleculares.

Formar la base de la limpieza de los datos biológicos.

Errores, discrepancias, redundancias, ambigüedades y datos incompletos reducen la calidad de los datos en las bases de datos moleculares.

Fuentes de artefactos de los datos biológicos

1. Diversas fuentes de datos
 - a. Amplia duplicación
 - b. Reiteradas peticiones de las secuencias a la misma o a distintas bases de datos
 - c. Actualización cruzada de las bases de datos (propagación de errores)
2. Anotación de los datos

- a. Enriquecimiento de las secuencias con las descripciones de sus características estructurales y funcionales y otras referencias relacionadas con la información secuencial.
 - b. Mediante los administradores de las bases de datos o los investigadores que suben las secuencias.
 - c. Las bases de datos tienen diferentes mecanismos para la anotación de los datos (GENBANK – sólo la carga directa, SWISS-PROT – secuencia de todos los registros)
 - d. Errores de entrada de los datos se pueden llevar a diferentes interpretaciones.
3. Falta de normalización en la nomenclatura
 - a. Variaciones de convenciones en la nomenclatura.
 - b. Sinónimos, homónimos y abreviaturas.
 4. Insuficiencia de mecanismos de control de la calidad de los datos
 - a. Enfoques sistemáticos para la depuración de los datos son insuficientes.

N.4. Clasificación de los artefactos de los datos

Cabecera: Información general del registro.

Característica: Descripción de lo estructural, funcional y otras propiedades físico – químicas de la secuencia y las regiones de interés.

Secuencia: secuencia de nucleótidos o proteínas.

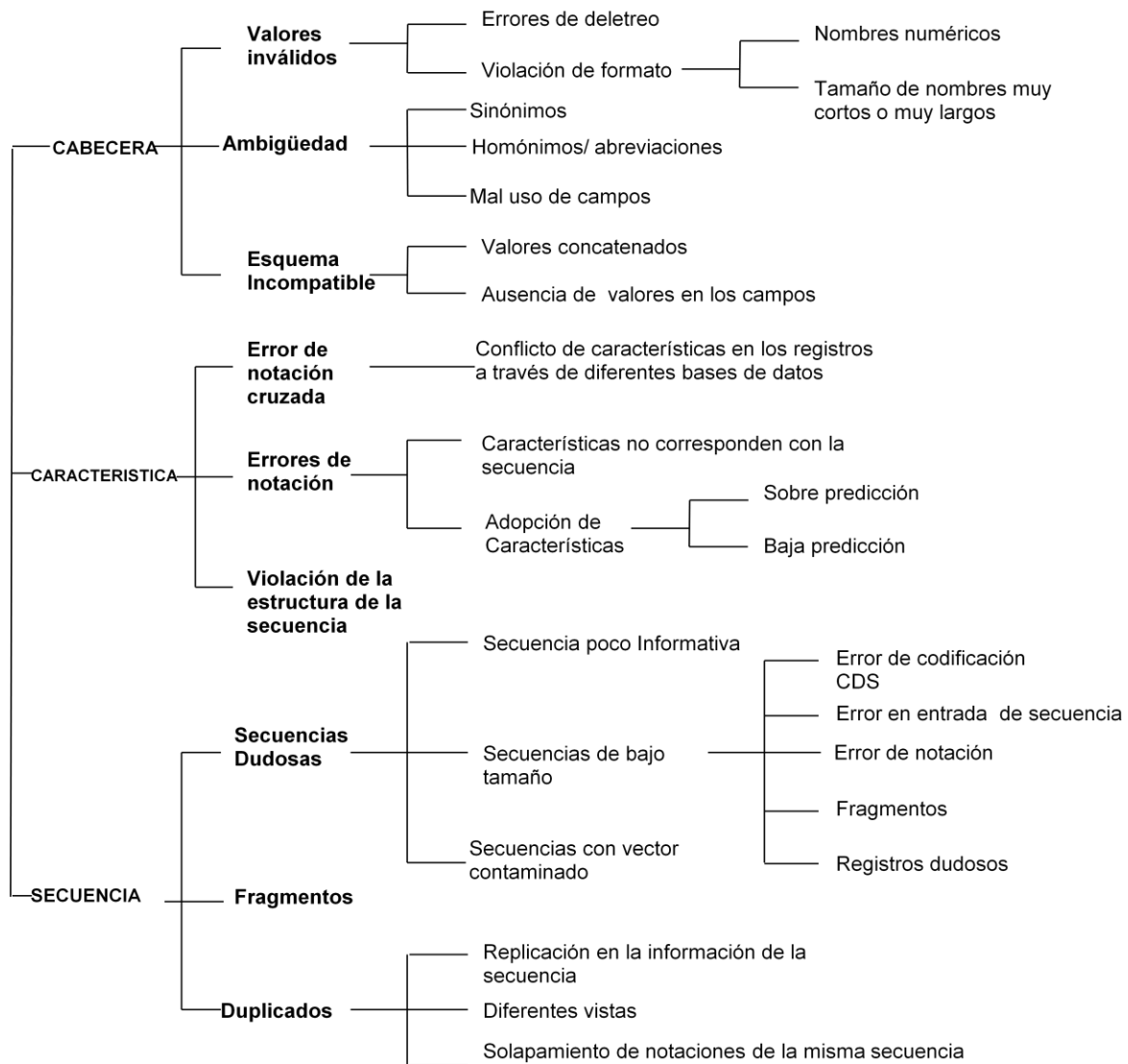


Figura N1. Errores 1

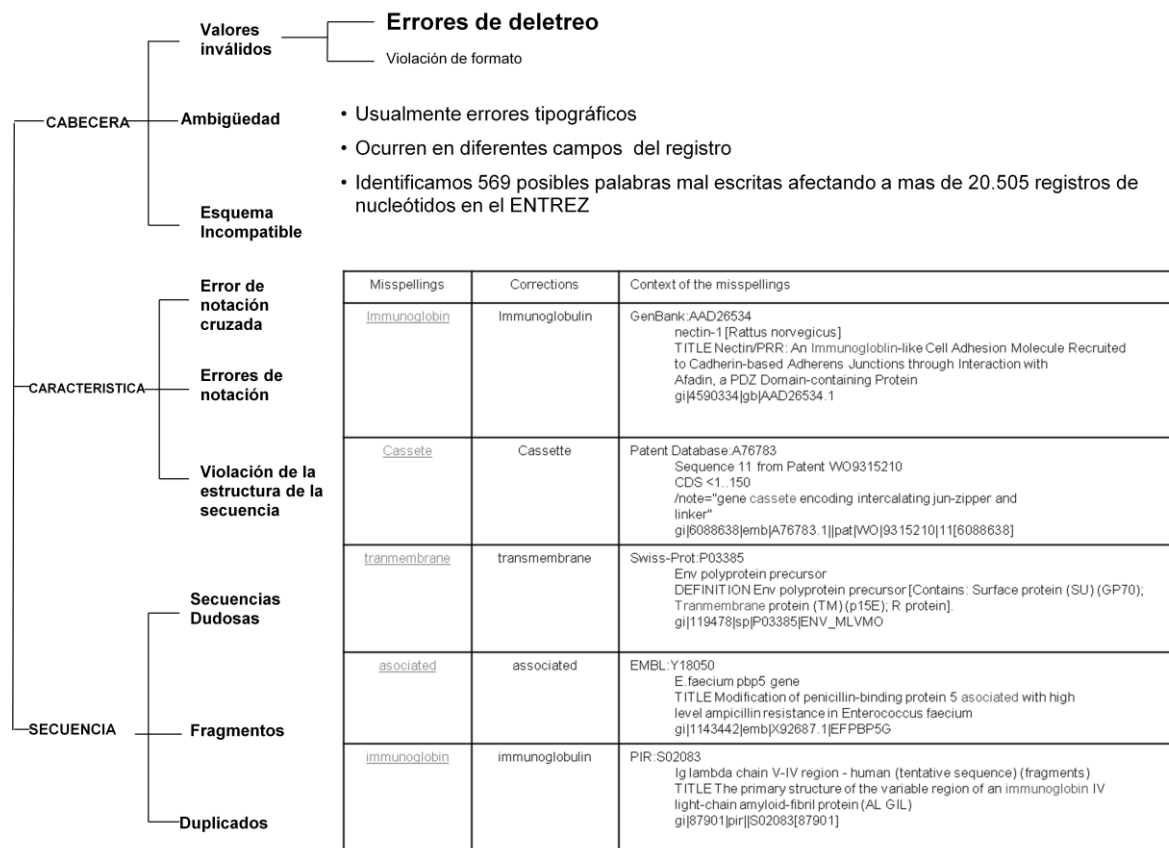


Figura N2. Errores 2

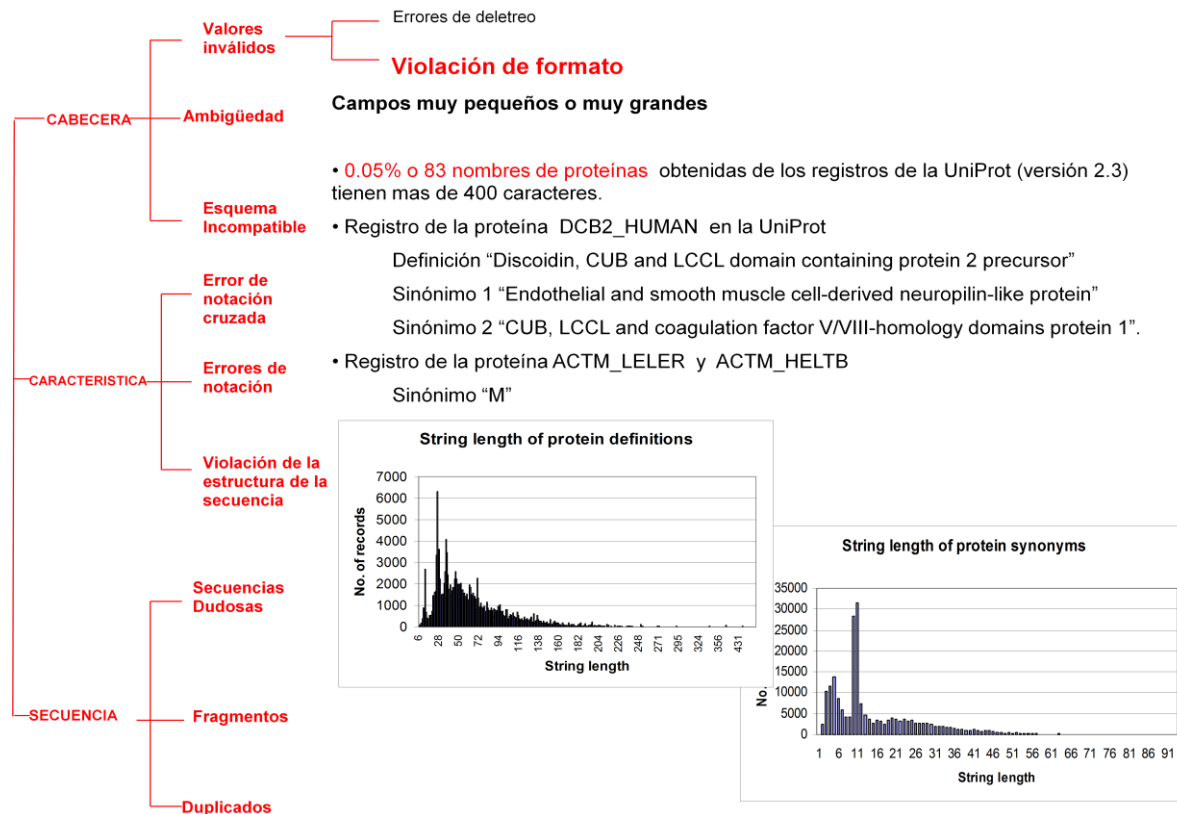


Figura N3. Errores 3

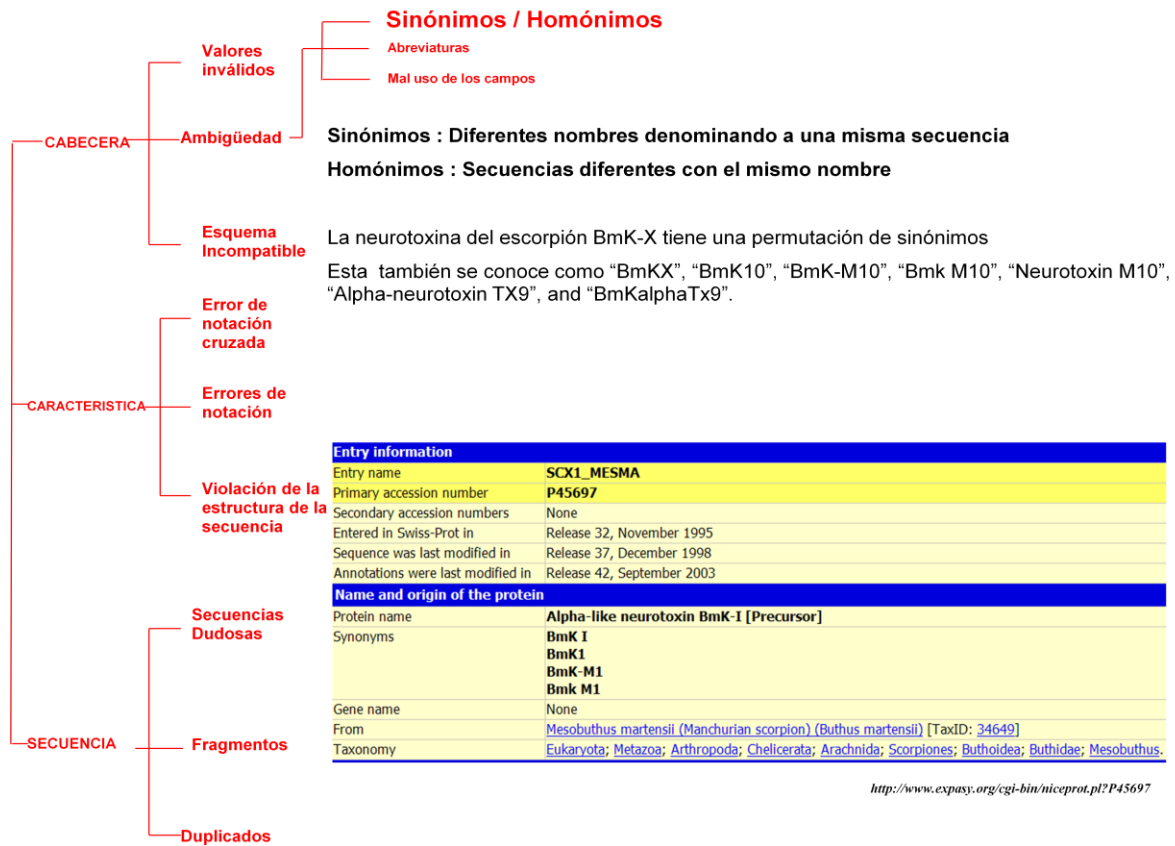


Figura N4. Errores 4

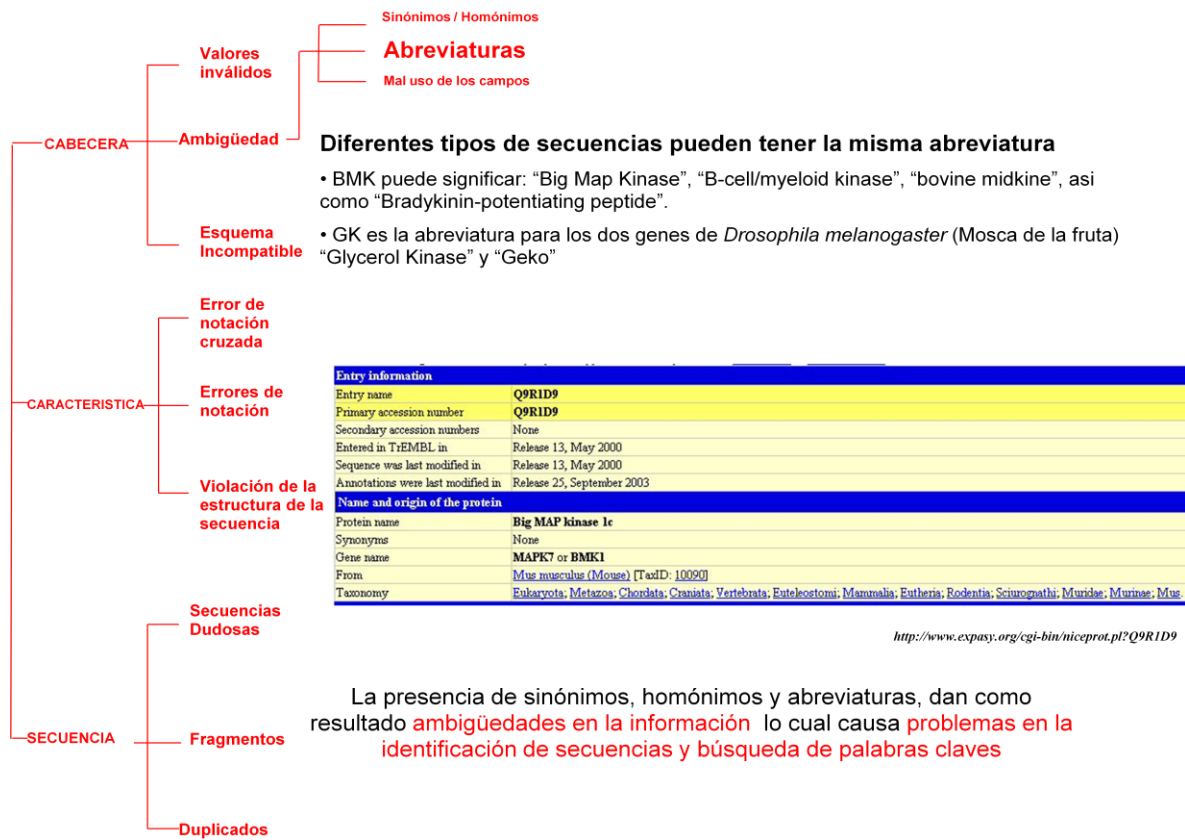


Figura N5. Errores 5



Figura N6. Errores 6

Valores concatenados
 Ausencia de valores en los campos

La concatenación de los campos ocurre durante la transformación de los datos

- Cuando los campos de granularidad mas fina son transformados dentro del esquema a sus correspondientes campos de granularidad mas gruesa, los valores de los campos son concatenados
- Múltiples valores de los campos pueden ser concatenados usando “and” o “or”.

El nombre del gen, de entrada P29834 Swiss-Prot fue “GRP 0.9 or GRP-1”. Esto fue corregido recientemente.

CABECERA

- Valores inválidos
- Ambigüedad
- Esquema Incompatible

CARACTERISTICA

- Error de notación cruzada
- Errores de notación
- Violación de la estructura de la secuencia

SECUCENCIA

- Secuencias Dudosas
- Fragmentos
- Duplicados

Entry information	
Entry name	AEP_MESMA
Primary accession number	P15228
Secondary accession numbers	None
Entered in Swiss-Prot in	Release 14, April 1990
Sequence was last modified in	Release 14, April 1990
Annotations were last modified in	Release 41, February 2003
Name and origin of the protein	
Protein name	Anti-epilepsy peptide [Fragment]
Synonym	AEP
Gene name	None
From	Mesobuthus martensii (Manchurian scorpion) (Buthus martensii) [TaxID: 34649]
Taxonomy	Eukaryota ; Metazoa ; Arthropoda ; Chelicerata ; Arachnida ; Scorpiones ; Buthoidea ; Buthidae ; Mesobuthus
References	
[1]	SEQUENCE. TISSUE= Venom ; MEDLINE=89193444; PubMed=2930463; [NCBI, ExPASy, EBI, Israel, Japan] Zhou X.-H. , Yang D. , Zhang J.-H. , Liu C.-M. , Lei K.-J. ; "Purification and N-terminal partial sequence of anti-epilepsy peptide from venom of the scorpion <i>Buthus martensii</i> Karsch.", Biochem. J. 257:509-517(1989)
Comments	
	<ul style="list-style-type: none"> • FUNCTION: Shows anti-epileptic activity. • SUBCELLULAR LOCATION: Secreted. • TISSUE SPECIFICITY: Expressed by the venom gland. • SIMILARITY: Belongs to the alpha/beta-scorpion toxin family. Alpha-toxin subfamily.

<http://www.expasy.org/cgi-bin/niceprot.pl?P15228>

Figura N7. Errores 7

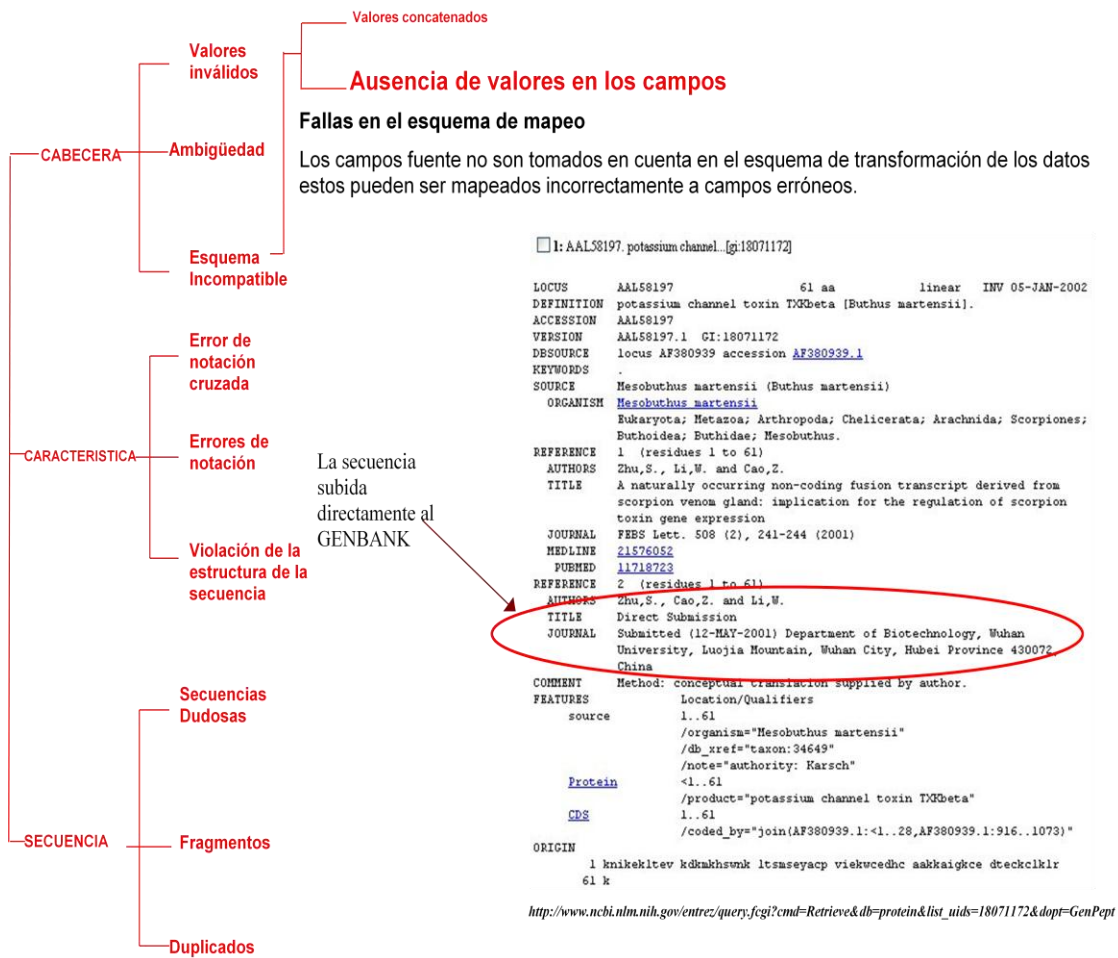


Figura N8. Errores 8

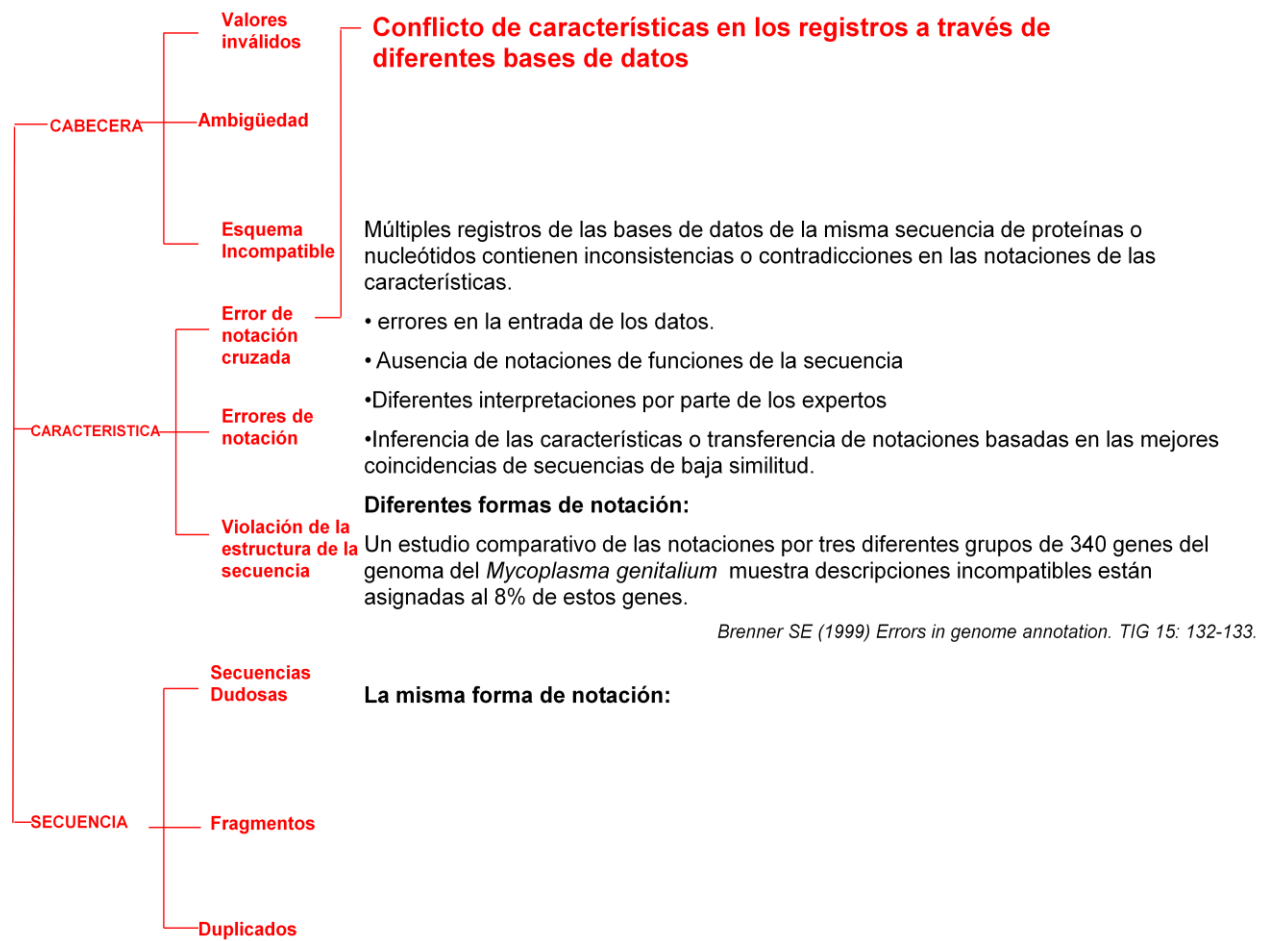


Figura N9. Errores 9

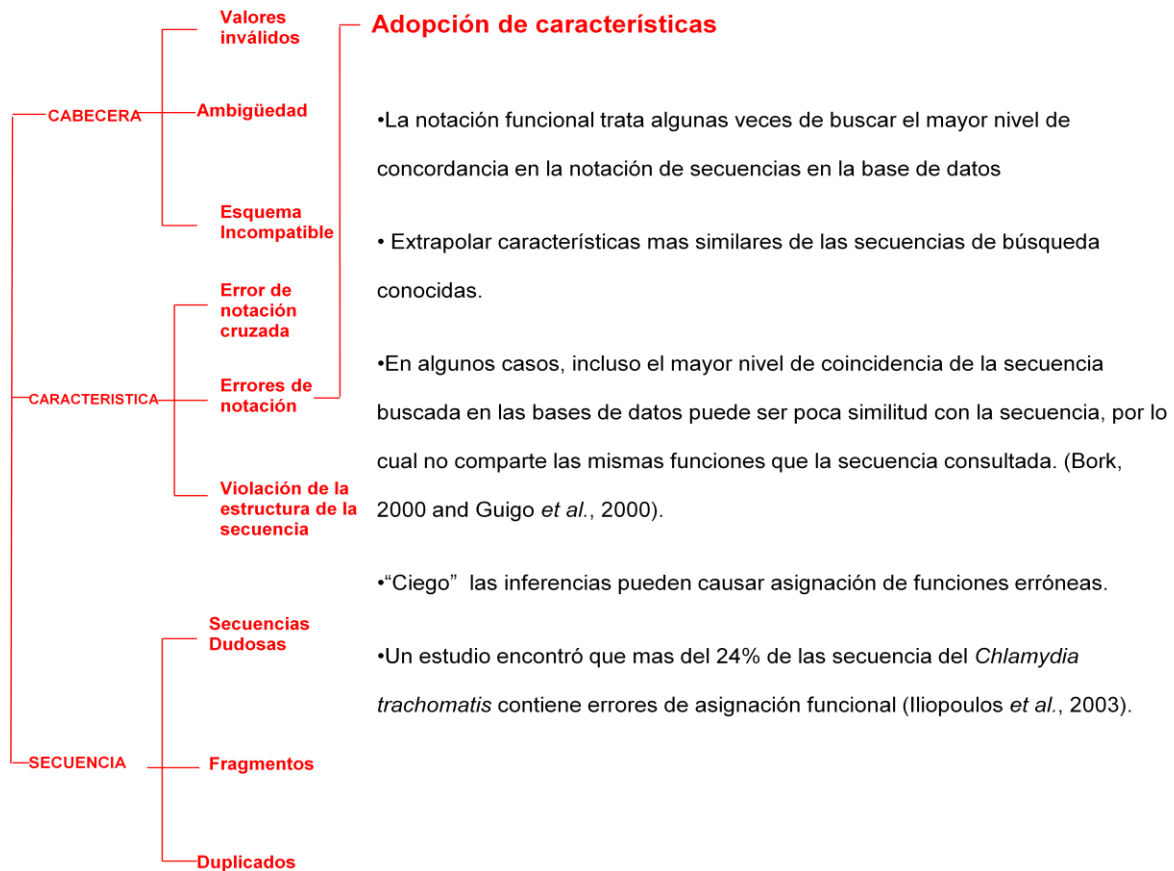


Figura N10. Errores 10

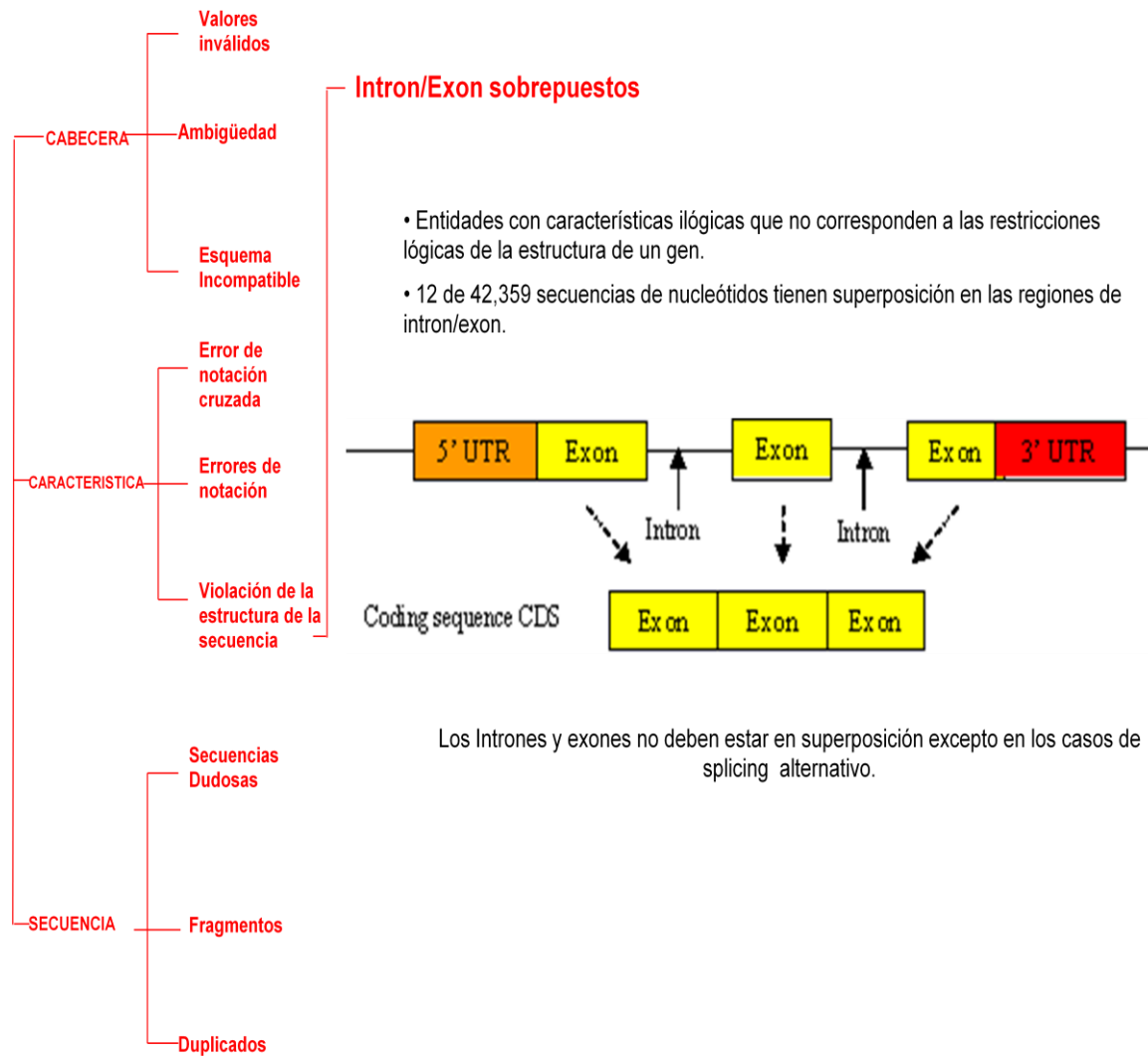


Figura N11. Errores 11

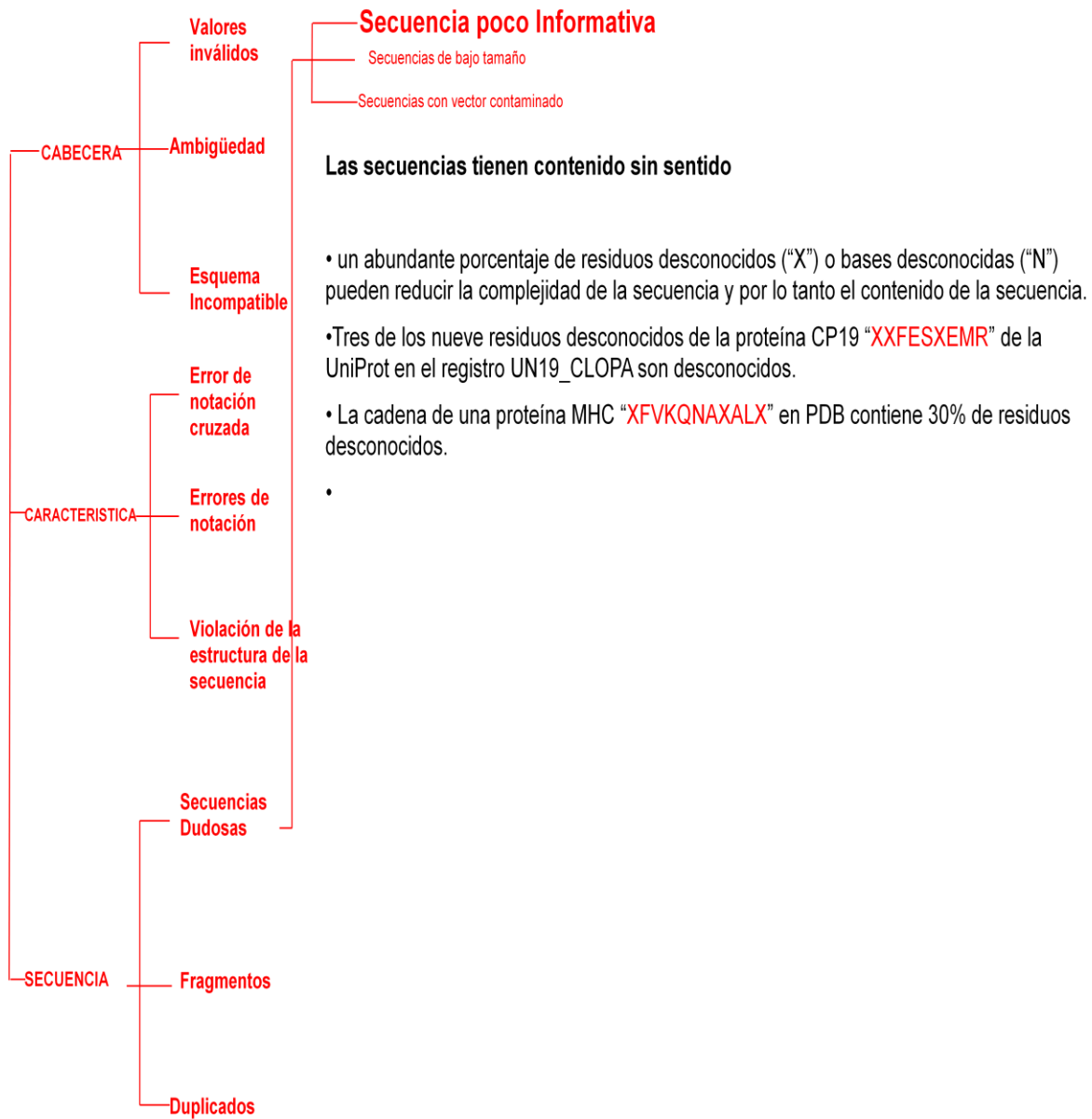


Figura N12. Errores 12

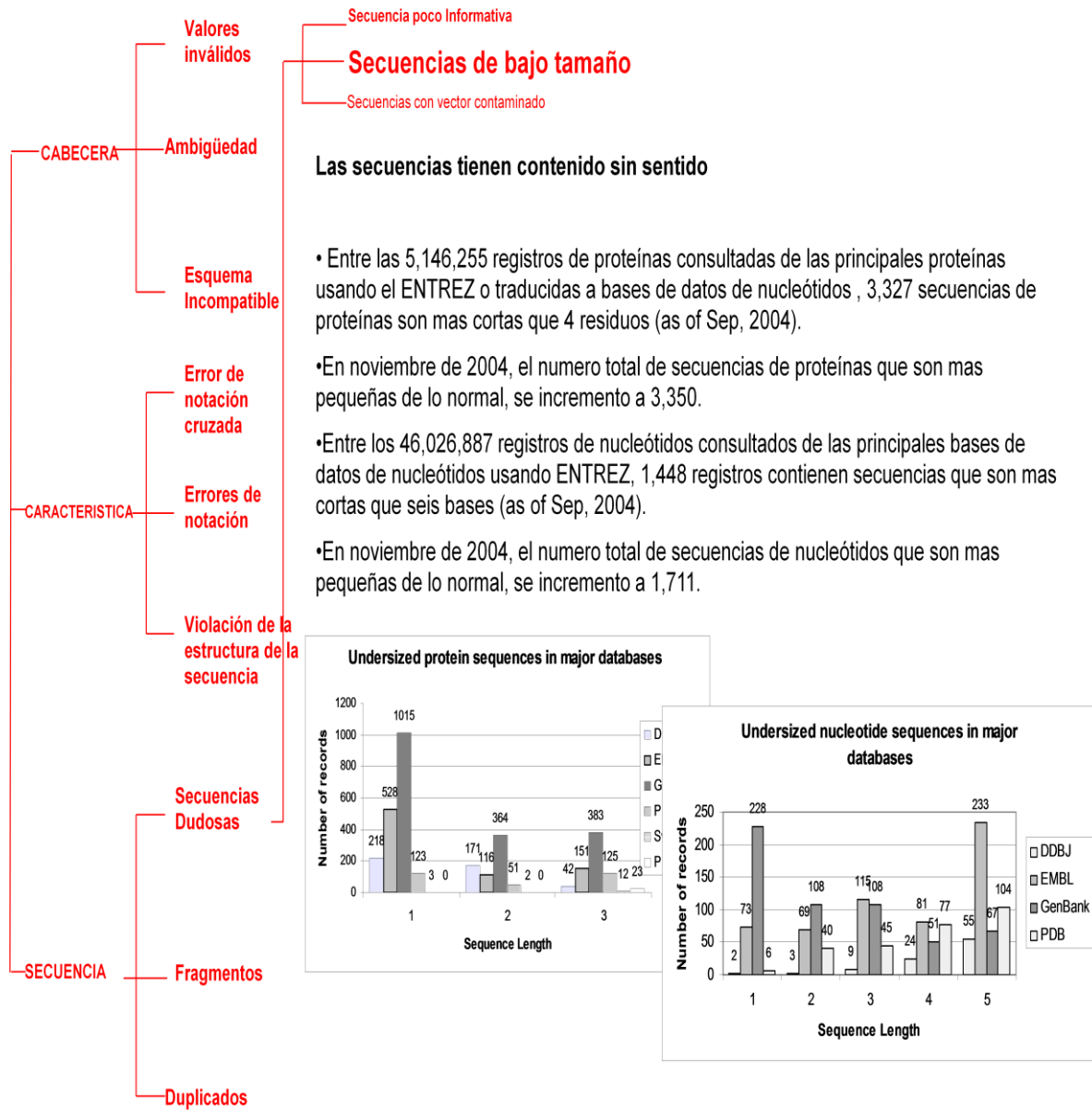


Figura N13. Errores 13

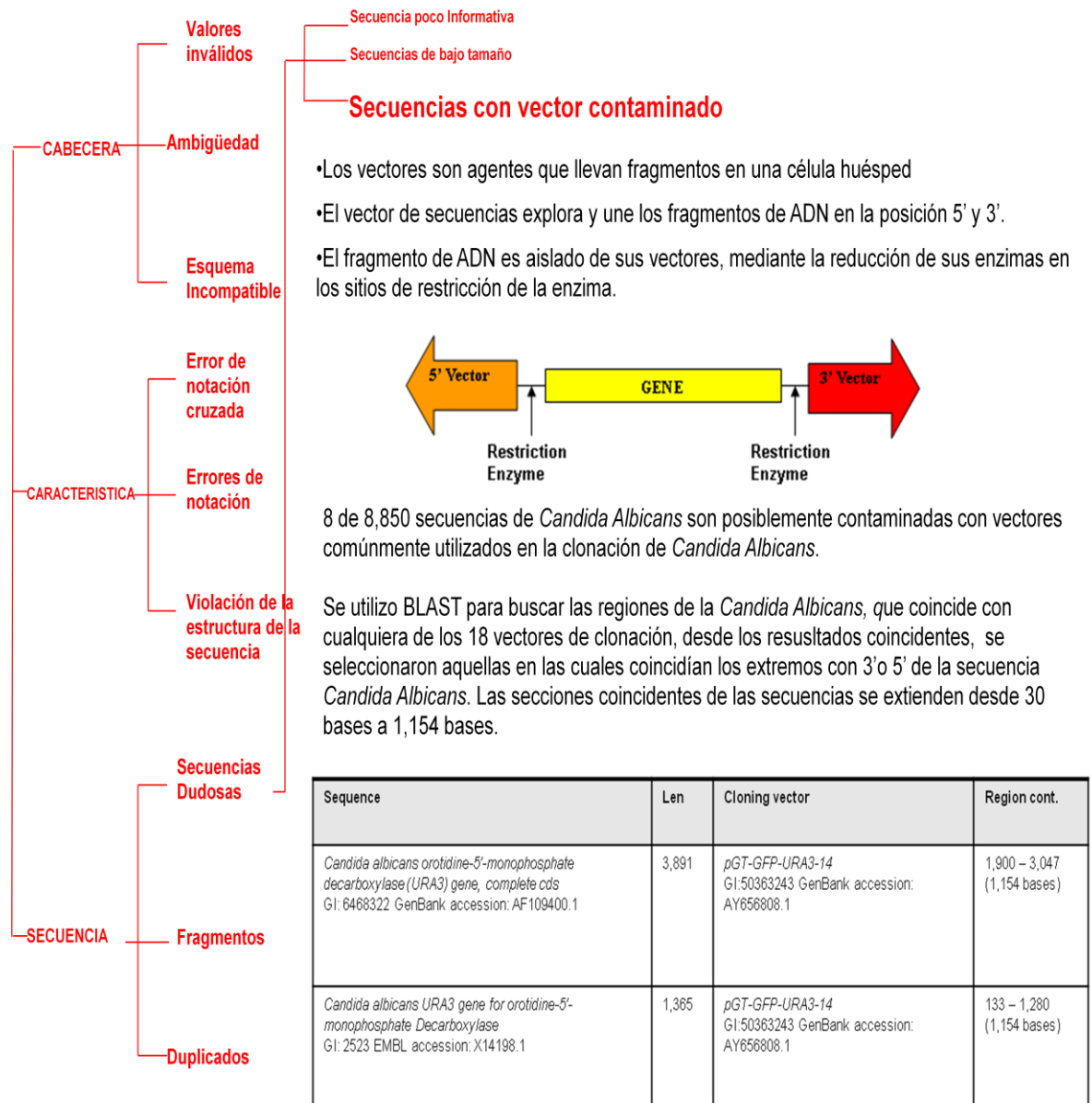


Figura N14. Errores 14

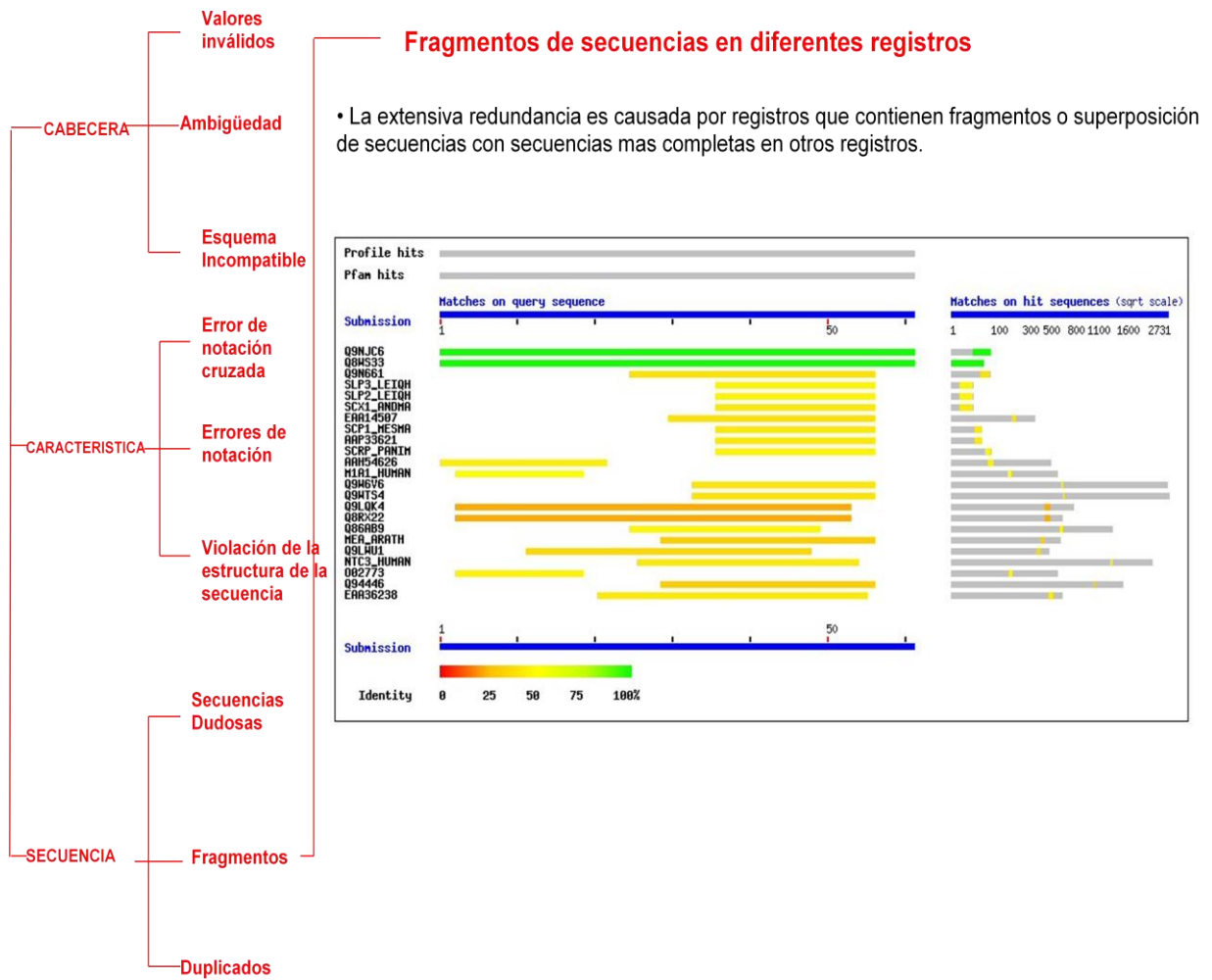


Figura N15. Errores 15

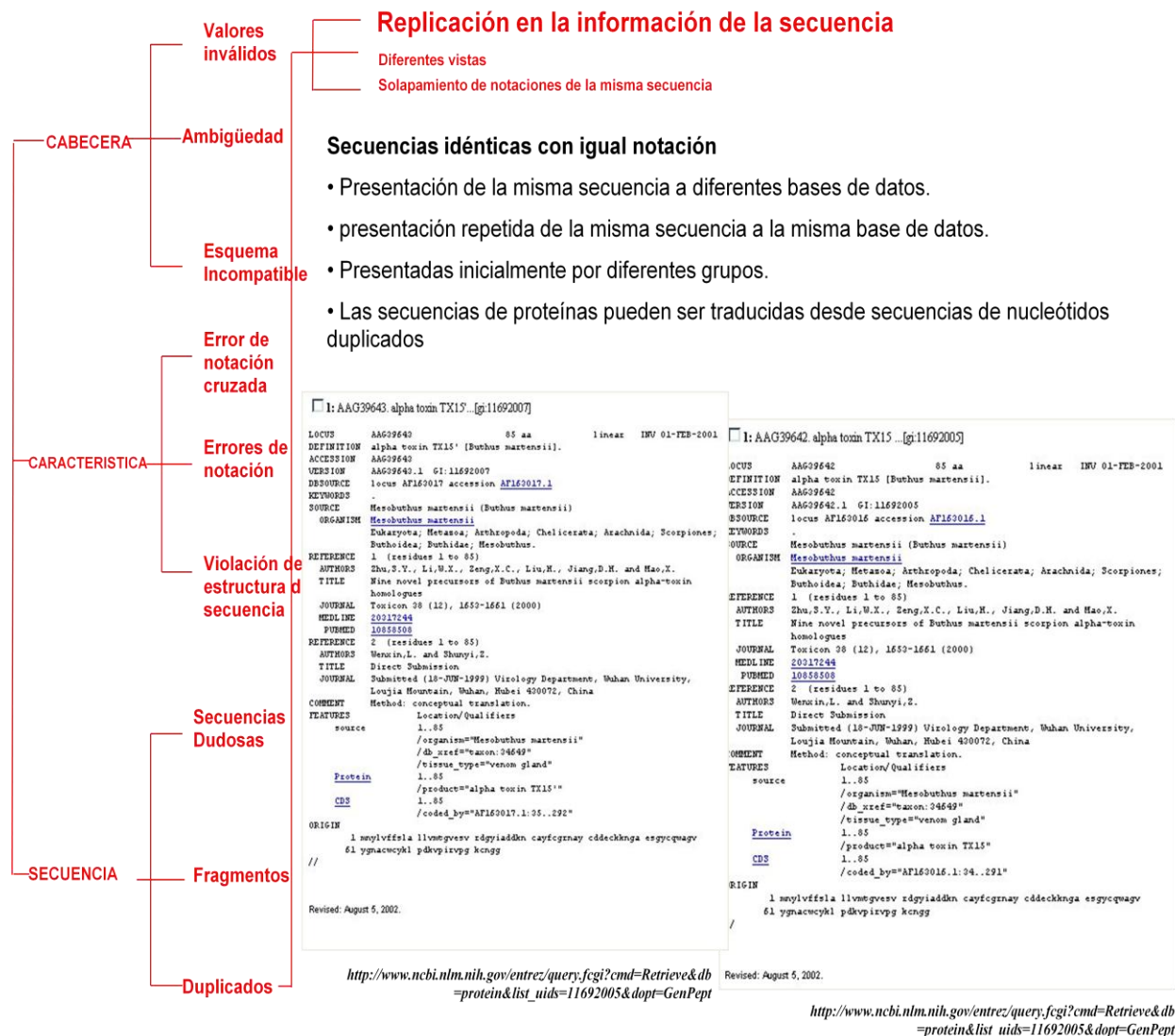


Figura N16. Errores 16

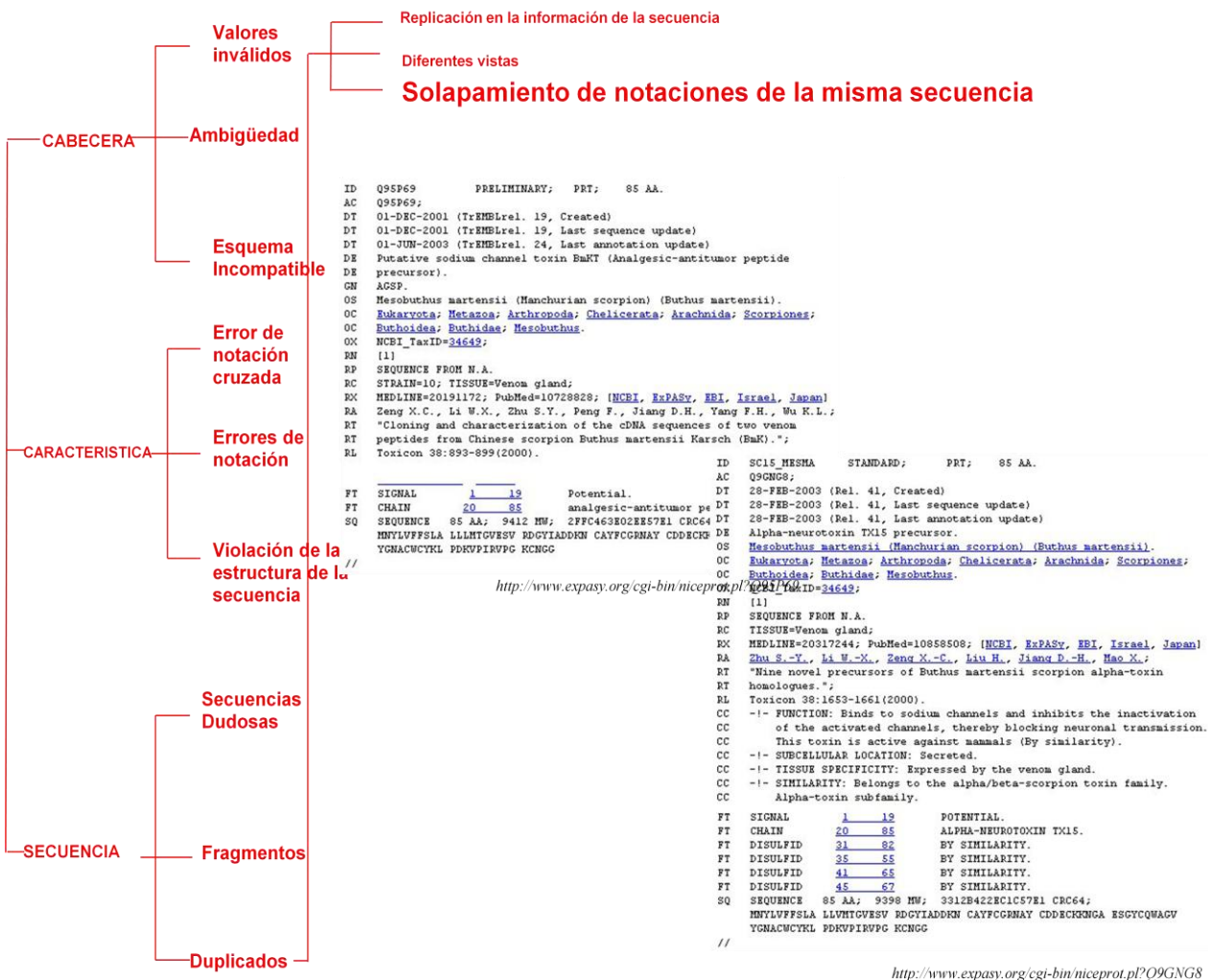


Figura N17. Errores 17

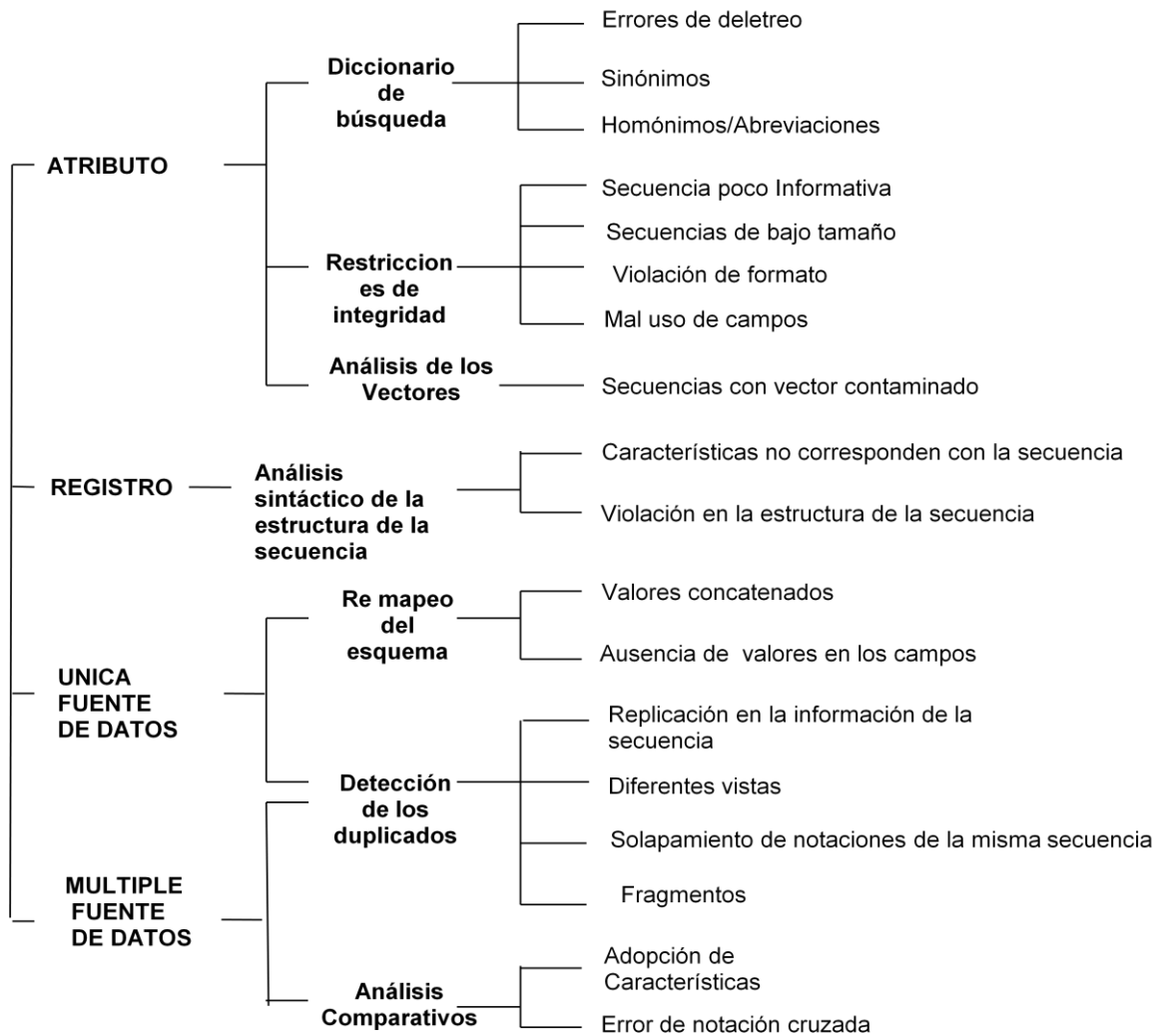


Figura N18. Errores 18