

# Universidad de las Ciencias Informáticas

## Facultad 6



### **Título: Diseño de una aplicación para el análisis de imágenes en la caracterización genética de microorganismos.**

Trabajo de Diploma para optar por el título de  
Ingeniero en Ciencias Informáticas

**Autor(es):** Sianny Socorro Suarez

Rigoberto Miguel Alvarez Pérez.

**Tutor (es):** Dr. Raúl Díaz.

Lic. Orlando Martínez.

Ing. Maypher Duran.

**Co-tutor:** Lic. Sandy Henríquez Villafruela.

**Ciudad de La Habana, Julio 2007**

## **DECLARACIÓN DE AUTORÍA**

Declaramos ser autores de la presente tesis y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmo la presente a los \_\_\_\_ días del mes de \_\_\_\_\_ del año \_\_\_\_\_.

**<nombre autor>**

**<nombre tutor>**

\_\_\_\_\_

\_\_\_\_\_

Firma del Autor

Firma del Tutor

## **OPINIÓN DEL TUTOR DEL TRABAJO DE DIPLOMA**

**Título: Diseño de una aplicación para el análisis de imágenes en la caracterización genética de microorganismos.**

Autor: Sianny Socorro Suárez

Rigoberto Miguel Álvarez Pérez.

El tutor del presente Trabajo de Diploma considera que durante su ejecución los estudiantes mostraron las cualidades que a continuación se detallan:

Los estudiantes comenzaron a trabajar en el proyecto de análisis de imágenes en la caracterización genética de microorganismos desde etapas tempranas de su concepción, mostrando gran interés y preocupación por el desarrollo del proyecto. Es de destacar la disposición en la preparación autodidacta de herramientas de desarrollo de software que fue necesario estudiar. La comunicación con el tutor fue muy positiva y constante, además los estudiantes manifestaron un interés continuo en aspectos relacionados con el desarrollo de su proyecto, acudiendo a su tutor en todo momento. El intercambio de conocimientos fue recíproco. Ambos compañeros se caracterizan por un sentido de responsabilidad muy elevado. A pesar de los obstáculos que enfrentaron en los inicios del proyecto, por lo novedoso del tema, eso no los desanimó y continuaron interesándose por el desarrollo de su trabajo de tesis. La solución informática propuesta no es más que una aplicación para el análisis de imágenes en la caracterización genética de microorganismos para el Centro de Medicina Tropical Pedro Kourí, el cual pretende dar solución a un problema actual en dicha organización, y para lograr eso fue necesario un estudio profundo de algunas tecnologías.

Por todo lo anteriormente expresado considero que los estudiantes están aptos para ejercer como Ingenieros Informáticos; y propongo que se le otorgue al Trabajo de Diploma la calificación de Excelente. Además, considero que los resultados poseen valor para ser publicados.

**La ciencia se compone de errores, que a su vez son los pasos hacia la verdad.**

**Julio Verne.**



## **Agradecimientos.**

Son muchas las personas que han apoyado a la realización de este trabajo y a las que no queremos perder la oportunidad de agradecer por su esfuerzo y ayuda incondicional.

A Julio Omar por brindar su conocimiento sin límites, por estar dispuesto cada vez que lo necesitamos.

A Lisbell por su preocupación, y su ayuda desmedida.

A Danay y Daniel por siempre darnos buenas ideas y brindarnos su conocimiento.

A Yudith por ayudarnos con la ingeniería.

A Andy por darnos apoyo y ayuda.

Agradecemos además a la Revolución y a nuestro Comandante en Jefe, Fidel Castro, por hacernos parte de este, su gran sueño.

A todos nuestros profesores, por transmitirnos los conocimientos y valores que nos permitieron formarnos como profesionales.

A nuestros tutores Orlando Martínez, Maypher Duran y Raúl Díaz por brindarnos sus conocimientos, ser atentos y preocupados.

A Sandy Henriquez por ayudarnos con el documento y por su preocupación.

A todas nuestras amistades por su compañía durante todos estos años, por compartir los buenos y malos momentos.

Y de manera muy especial a nuestros padres por su dedicación, confianza y amor. Por guiarnos por el camino correcto y confiar siempre en nosotros.

A todos muchas GRACIAS.

**Dedicatoria:**

De Sianny:

Por el inmenso respeto y amor que les tengo, dedico este trabajo a mis padres Ernesto y Lissette, por haberme dado la oportunidad de estar en este mundo, por darme tanto amor. Por haber hecho de mi lo que soy, con su ejemplo y sabiduría, por siempre tenerme confianza y apoyarme, por enseñarme valores y principios que me han ayudado a lo largo de la vida. Sin ustedes no hubiese logrado nada.

A mi hermano Sandy por ayudarme y quererme, a Eliete por ser la hermana que no tengo, a mis abuelos, por su ternura y confianza, porque son mis otros padres, los quiero mucho.

A mis tíos y primos por apoyarme.

A todas mis amistades de la universidad, en especial a las que han estado a mi lado los 5 años apoyándome en los malos momentos y compartiendo en los buenos, los llevo en mi corazón para siempre.

A Rayner por estar en las buenas y malas, por ser mi apoyo y compañía durante muchos años. A Clorys, Pepe, Fide, Ángel y Olguita por permitirme ser parte de sus vidas y ayudarme siempre.

A mis amistades de la infancia y del pre que han seguido mi carrera de cerca y me han apoyado.

A mi compañero de tesis Rigoberto sin el no hubiese logrado el éxito.

A todas las personas que han formado parte de mi vida en este ultimo año, por darme buenos consejos y ayudarme.

A todos por su cariño y confianza GRACIAS !!!!

De Rigoberto:

Quisiera, de manera muy especial, dedicar este trabajo a mi mamá, por su apoyo incondicional en cada paso que doy en la vida, por permitir ser mi guía y ejemplo, por que me das las fuerzas para seguir adelante, por ser la fuente de mi inspiración. Te quiero mucho.

A Rubén por ser como un padre para mí. Quizás no nos unan lazos de sangre, pero nos une el cariño y el afecto.

A mi papá, por su amistad y preocupación.

A mi abuelita, que aunque ya no se encuentre entre nosotros, soñó con verme este día

A mi hermana Zulema, y mi sobrina Zuyamis por ser tan especiales.

A mis amigos incomparables, Andy Yaikiel, Duniel, Julio, Yaiser, Yudel, Joel

A Lisbell, que me soporta cada día. Su ayuda incalculable ha sido un sostén para mi éxito.

A Sianny, mi compañera de tesis, por su apoyo y dedicación.

A todas las personas que de una forma u otra, contribuyeron a mi formación profesional y compartieron a mi lado durante estos 5 años, gracias.

## Resumen

El análisis de cluster sobre imágenes de caracterización genética tiene gran aplicación en diferentes campos de investigación: forense, mercadeo, filogenia, epidemiología, etc. Sin embargo, existen importantes centros en nuestra isla que no pueden acceder a los software que realicen dicho análisis por los costes que representa.

El trabajo surge por la necesidad de contar en el Instituto de Medicina Tropical “Pedro Kourí” (IPK) con una aplicación que realice el análisis de conglomerados (clusters) de muestras de ADN en la caracterización genética de microorganismos. Con este trabajo se pretendió implementar una aplicación que permita el análisis de cluster de imágenes digitales de muestras de caracterización genética y luego visualice gráficamente los cluster resultantes de dicho análisis. Como resultado de la investigación se planteó como objetivo diseñar una aplicación para automatizar el análisis de cluster de imágenes digitales de resultados de técnicas de caracterización genética. La misma permite principalmente que se realice de manera automatizada, rápida y con menos probabilidad de errores el trabajo de los especialistas o científicos que trabajan con estas técnicas. Además de disminuir los costos que implicaría la compra de un software de este tipo ha dicho centro. Para garantizar la portabilidad de la aplicación el diseño se realizó con la metodología RUP y como herramienta el Rational Rose.



# Índice

<b>INTRODUCCIÓN .....</b>	<b>1</b>
<b>CAPÍTULO 1 FUNDAMENTACIÓN TEÓRICA .....</b>	<b>4</b>
1.1 INTRODUCCIÓN .....	4
1.2 CONCEPTOS ASOCIADOS AL DOMINIO DEL PROBLEMA .....	4
1.2.1 Polimorfismo .....	4
1.2.2 Marcadores moleculares .....	5
1.2.3 Análisis de Clusters.....	8
1.3 ESTADO DEL ARTE .....	8
1.3.1 GelCompar.....	9
1.3.2 BioNumeric.....	9
1.3.3 GelQuant .....	10
1.3.4 TotalLab.....	10
1.3.5 Mat-Gen .....	10
¿POR QUÉ DESARROLLAR UNA APLICACIÓN NACIONAL? .....	10
1.4 METODOLOGÍAS QUE PUEDEN UTILIZARSE PARA EL DESARROLLO DEL SISTEMA. ....	11
1.4.1 RUP.....	11
1.4.2 Metodología XP (Extreme Programming).....	13
1.4.3 Metodología FDD (Desarrollo guiado por la funcionalidad, Feature driven development en ingles).....	14
1.5 HERRAMIENTAS UTILIZADAS PARA EL DESARROLLO DEL SISTEMA. ....	14
1.5.1 Rational Rose Enterprise Edition.....	15
1.6.2 Visual paradigma – UML.....	16
1.6 ROLES DE RUP DESEMPEÑADOS EN EL DESARROLLO DE LA APLICACIÓN: .....	16
1.7 CONCLUSIONES PARCIALES: .....	20
<b>CAPÍTULO 2 CARACTERÍSTICAS DEL SISTEMA.....</b>	<b>21</b>
2.1 INTRODUCCIÓN .....	21
2.2 MODELO DEL DOMINIO. ....	21
2.2.1 Definición de las entidades y los conceptos principales.....	21
2.2.2 Diagrama del modelo de dominio.....	23
2.2.3 Reglas del negocio .....	23
2.3 ESPECIFICACIÓN DE LOS REQUISITOS DEL SOFTWARE .....	24
2.3.1 Requerimientos Funcionales.....	24
2.3.2 Requerimientos No Funcionales. ....	25
2.4 DIAGRAMA DE CASOS DE USO DEL SISTEMA .....	27
2.4.1 Descripción de los casos de uso.....	27
2.4.2 Definición de los actores. ....	28

---

2.4.3 Listado de casos de uso .....	28
2.5 PROTOTIPOS DE INTERFAZ DE USUARIO. ....	30
2.6 CONCLUSIONES PARCIALES .....	32
<b>CAPÍTULO 3 ANÁLISIS Y DISEÑO DEL SISTEMA .....</b>	<b>33</b>
3.1 INTRODUCCIÓN. ....	33
3.2 ANÁLISIS .....	33
3.2.1 Modelo conceptual de clases de análisis.....	33
3.3 DIAGRAMA DE CLASES DE ANÁLISIS .....	34
3.4 PRINCIPIOS DE DISEÑO.....	34
3.5 PATRONES .....	35
3.5.1 Patrones de arquitectura.....	35
3.5.2 Patrones de diseño.....	35
3.6 DISEÑO.....	36
3.6.1 Diagrama de Clases del Diseño.....	36
3.6.2 Descripción de las clases. ....	36
3.6.2 Diagramas de Interacción.....	45
3.6.2.1 Diagrama de Colaboración.....	45
3.6.2.2 Diagrama de Secuencia.....	46
3.7 DISEÑO DE LA BASE DE DATOS. ....	46
3.8 MODELO DE DESPLIEGUE .....	47
3.9 GENERALIDADES DE LA IMPLEMENTACIÓN .....	48
3.10 CONCLUSIONES PARCIALES.....	49
<b>CONCLUSIONES.....</b>	<b>50</b>
<b>RECOMENDACIONES. ....</b>	<b>51</b>
<b>BIBLIOGRAFÍA .....</b>	<b>52</b>
<b>ANEXOS .....</b>	<b>55</b>
<b>GLOSARIO DE TÉRMINOS.....</b>	<b>92</b>

### Introducción

Un lunes 9 de septiembre de 1984 el profesor Alec Jeffreys sacó del tanque de revelado, en su laboratorio G18, en la planta baja del departamento de Genética de la Universidad de Leicester (Reino Unido), una fotografía que mostraba una imagen parecida a un código de barras. Aquel día se había inventado, la huella digital genética (DNA fingerprinting, en inglés), uno de los descubrimientos científicos que ha tenido mayor impacto en la vida de millones de seres humanos.

La fotografía obtenida aquel 9 de septiembre por el profesor Jeffreys se conseguía mediante la extracción de material de la saliva o el pelo de una persona, y aquella imagen tan parecida a un código de barras era única para cada individuo con un porcentaje abrumador de fiabilidad. [1]

Fue en 1985 cuando se afirma que existe en el genoma humano, regiones de gran variabilidad, que permiten identificar un individuo de otro. [2]El análisis de estas variaciones, sobre imágenes de caracterización genética, permite identificar patrones para agrupar las muestras por comportamientos semejantes utilizando técnicas de agrupamientos o conglomerados (clusters en inglés). Sus aplicaciones pueden llegar desde diagnóstico prenatal de desórdenes genéticos e investigación forense de restos de individuos, hasta el monitoreo de progresión de enfermedades infecciosas; por solo citar algunos ejemplos.

Cuba es uno de los países que más atención le presta a la salud pública. En los últimos 20 años, se han realizado grandes esfuerzos para crear centros de excelencia médica que ayuden al control y eliminación de enfermedades que afectan nuestra población. Uno de estos centros es el Instituto de Medicina Tropical "Pedro Kourí" (IPK), que vierte sus esfuerzos desde su creación, en el diagnóstico, control e investigación de enfermedades infecciosas.

Dotados con las más modernas técnicas y con un grupo de reconocidos especialistas, el IPK tiene como principales objetivos, la protección de nuestra población contra las enfermedades tropicales nativas y la prevención de la introducción de enfermedades exóticas, la colaboración en la lucha contra las enfermedades tropicales en los países subdesarrollados, contribuir al desarrollo de la ciencia biomédica y la medicina tropical en particular, y la atención integral de calidad al paciente del Síndrome de Inmunodeficiencia adquirida (SIDA) y a otras enfermedades transmisibles.

Para complementar sus objetivos, el IPK utiliza, entre otras, técnicas de caracterización genética de microorganismos para identificar comportamientos semejantes entre las muestras, que permiten definir

agrupamientos para diferenciar las cepas en base a sus características genéticas (taxonomía). Este proceso tiene implicaciones muy importantes para el control de enfermedades permitiendo definir cadenas de transmisión. Sin embargo, el análisis de cluster, es actualmente realizado de manera visual en el centro, trayendo consigo un inconveniente a tener en cuenta: cuando se trata de pocas muestras, el procedimiento se puede realizar sin mucho problema, pero a la hora de realizarlo a una cantidad mayor de muestras, se pueden cometer errores en el momento de definir los clusters.

En Cuba no se cuenta con una aplicación confiable que realice análisis de cluster de imágenes digitales de caracterización genética (huellas digitales genéticas). En el mercado internacional, el precio de estos productos oscila entre los 2000 y 10000 euros, cifras por las cuales el IPK no puede acceder a estos productos.

Este trabajo surge como necesidad de dar solución a las situaciones antes expuestas; por lo que nuestro **problema** consiste en ¿Cómo mejorar el proceso para realizar el análisis de cluster sobre imágenes digitales de caracterización genética en el Instituto de Medicina Tropical Pedro Kourí?

Del siguiente trabajo se espera, que fundamente las bases para la creación de un producto nacional de software y dar una propuesta del análisis y diseño. Como **objeto de estudio** se tiene el reconocimiento de patrones de las muestras digitales de caracterización genética. Y el **campo de acción** de la aplicación que se pretende desarrollar es el análisis de las imágenes de las muestras de caracterización genética.

Para conseguir el desarrollo satisfactorio de la aplicación se tiene el siguiente **objetivo general**:

Diseñar una aplicación para automatizar el análisis de cluster de imágenes digitales del resultado de técnicas de caracterización genética.

Como **objetivos específicos**:

Gestionar imágenes digitales de muestras de caracterización genética de microorganismos.

Análisis del modelo de negocio, captura de los Requisitos y confección de los Casos de Usos del Sistema.

Diseño de la aplicación y de la base de datos.

Para alcanzar dichos objetivos se planteó desarrollar las siguientes tareas:

Revisión de los temas relacionados con los diferentes sistemas de análisis de cluster y las principales tecnologías utilizadas para ellos.

Realizar entrevistas para garantizar que el trabajo satisfaga los objetivos propuestos.

Estudiar y seleccionar las técnicas para el trabajo con imágenes de geles a través de aplicaciones informáticas.

Seleccionar la metodología de análisis y diseño informático que facilite la creación y garantice la calidad del sistema, así como su mantenimiento.

Para lograr una mejor comprensión el presente documento se estructura en 3 capítulos de contenidos donde se incluye todo lo relacionado con el trabajo investigativo realizado.

En el Capítulo 1: Fundamentación Teórica, se definen conceptos con el objetivo de mejorar el entendimiento del trabajo, se describen algunos software dedicados al análisis de cluster en el mundo. También se realiza un análisis a las tendencias y tecnologías actuales con el objetivo de seleccionar las herramientas para desarrollar un producto de calidad.

El Capítulo 2: Modelo del dominio, describe la solución propuesta utilizando los componentes del modelo del dominio de la metodología RUP. De este modelo se tendrán en cuenta la definición de las entidades y los conceptos principales, también se conocerán las reglas del negocio. Así como los requisitos, describe los Requisitos Funcionales y No Funcionales del sistema, y los Actores que intervienen. También se encuentra el Diagrama de Casos de Uso del Sistema y la descripción de cada uno.

El Capítulo 3: Descripción de la solución propuesta, realiza el diagrama de clases del análisis, diagramas de interacción y el diagrama de clases del diseño, así como el diagrama de clases persistentes y el modelo de datos.

# CAPÍTULO 1

## Fundamentación Teórica

### 1.1 Introducción

En este capítulo se aborda una serie de conceptos necesarios para el entendimiento del tema a tratar, el estado actual de desarrollo de las aplicaciones que realizan el análisis de clusters sobre muestras de ADN en el ámbito mundial, además de una búsqueda de las tecnologías en las que se apoya la futura creación del sistema.

Han sido implementados muchos sistemas que realizan análisis de conglomerados a muestras de caracterización genética, que de una forma u otra responden a las exigencias, cada día más crecientes, de calidad y rapidez en todo proceso investigativo; pero por la carencia de ellos en el IPK expondremos las tecnologías y metodología que se va a utilizar para la creación de un producto nacional.

### 1.2 Conceptos asociados al dominio del problema

#### 1.2.1 Polimorfismo

El genoma humano es grandemente afectado por las variaciones de su secuencia (polimorfismo). Si tomamos dos individuos al azar y comparamos sus secuencias de ADN podremos comprobar que el 99,9% de sus secuencias son idénticas. En el 0,1% restante se encuentran las diferencias que determinan la susceptibilidad a enfermedades, la respuesta diversa al ambiente, el metabolismo diferencial ante drogas o medicamentos, así como un número importante de diferencias entre individuos. [3]

Los polimorfismos genéticos pueden ser identificados mediante digestión de una muestra de ADN con enzimas de restricción, lo que da como resultado fragmentos de restricción de longitud polimórfica o variable. [4]

Para distinguir uno de otro se separan los fragmentos mediante electroforesis (migración en un campo

eléctrico sobre una superficie de un gel) sobre geles de agarosa o poliacrilamida capaz de distinguir los fragmentos en función de sus diferentes tamaños. Luego estos fragmentos pueden visualizarse usando detección por radiactividad, fluorescencia o coloración con sales de plata. [5]

Desde hace casi 40 años, hay una tendencia a realizar la electroforesis usando marcadores moleculares, para la identificación de diferencias entre individuos.

### 1.2.2 Marcadores moleculares

Una serie de técnicas moleculares de gran desarrollo en los últimos veinte años permiten conocer la información genética que los organismos portan. Funcionan como señaladores de diferentes regiones del genoma y se los conoce en forma genérica como **marcadores moleculares**. Son ampliamente utilizados en genética humana, vegetal, animal y microbiana. Que permiten evidenciar variaciones (**polimorfismos**) en la secuencia del ADN entre dos individuos. [6]

Los primeros marcadores desarrollados a finales de los 70 se basaron en la identificación de proteínas e **isoenzimas** por electroforesis en geles de almidón o poliacrilamida. Con ellos se abrió el conocimiento de la estructura y heterogeneidad genética entre diferentes especies, variedades, y poblaciones de distinto origen geográfico. Pero esta técnica tenía una limitación muy importante: no era capaz de detectar suficiente polimorfismo entre variedades o especies próximas debido a que las proteínas son el resultado de la expresión génica, que puede ser distinta de unos tejidos a otros, de una etapa de desarrollo a otra, de un medio ambiente a otro, y de una época del año a otra. Los avances de la tecnología del ADN recombinante han permitido el desarrollo de los **marcadores moleculares basados en el ADN**, consiguiendo estabilidad en la identificación de especies y variedades.

Un marcador de ADN es simplemente un punto de referencia en un cromosoma, que puede o no corresponder a un gen. Diversas técnicas de biología molecular se encuentran disponibles para detectar variabilidad en la secuencia de ADN. La reacción en cadena de la polimerasa (PCR), las enzimas de restricción, la separación electroforética de los fragmentos de ADN, las sondas marcadas y las hibridaciones son algunas de las técnicas que permiten obtener un número virtualmente ilimitado de marcadores moleculares y cubrir la totalidad del genoma de un organismo. Cabe mencionar, que la técnica PCR, es esencial para la preparación de huellas digitales genéticas, si bien no vale para elaborar marcadores por sí sola [7].

Los marcadores moleculares se pueden agrupar en dos grandes grupos:

Basados en la identificación de proteínas:

- ✍ Isoenzimas

Basados en el ADN. Estos marcadores moleculares pueden ser clasificados a su vez en tres grupos, veamos algunos de cada tipo:

Marcadores basados en la hibridación del ADN:

- ✍ **RFLP «restriction fragment length polymorphisms»** o polimorfismos en la longitud de fragmentos de restricción de ADN.
- ✍ **Minisatélites o VNTR «variable number of tandem repeats»** son repeticiones en tandem de secuencias del genoma que contienen de 9 a 100 pares de bases.

Marcadores basados en la amplificación del ADN mediante la reacción de **PCR «polymerase chain reaction»** o reacción en cadena de la polimerasa de ADN:

- ✍ **RAPDs «random amplified polymorphic DNAs»** o fragmentos polimórficos de ADN amplificados aleatoriamente.
- ✍ **Microsatélites o SSR «simple sequence repeats»** o secuencias repetidas simples.

Marcadores mixtos:

- ✍ **AFLPs «amplified fragment length polymorphisms»** o polimorfismo en la longitud de fragmentos amplificados de ADN.
- ✍ **SCAR «sequence characterized amplified region»** o región amplificada de secuencia caracterizada. [8]

Las pruebas con marcadores moleculares, se han especializado y diversificado hacia otras aplicaciones de las ciencias médicas, a continuación intentamos una enumeración de las mismas:

En Investigación Genética:

- ✍ Diagnóstico prenatal de desórdenes genéticos.
- ✍ Diagnóstico de mutaciones
- ✍ Detección de portadores de genes específicos en familias y poblaciones.



- ✍ Generación de sondas para mapeo genético.
- ✍ Estudio de ligamentos de RFLP.
- ✍ Diagnóstico prenatal del sexo.
- ✍ Investigación forense de restos e individuos.

### En Investigación Básica en Biología Molecular:

- ✍ Síntesis de genes y modificación genética.
- ✍ Estudios de expresión génica.
- ✍ Secuenciación de ADN.

### En Investigación Clínica Médica:

- ✍ Detección y caracterización de patógenos.
- ✍ Identificación de oncogenes activados y caracterización de tumores.
- ✍ Monitoreo de progresión de enfermedades.
- ✍ Estudio de susceptibilidad y predisposición a enfermedades.

A manera de ejemplo de las aplicaciones médica, citamos un caso. El SIDA puede ser detectado amplificando una región de la proteína del virus y realizar la comparación con la muestra de la persona.

Tenemos las herramientas necesarias para poder identificar variaciones o polimorfismo entre un conjunto de muestras de ADN en el laboratorio. Dichas muestras pudieran pertenecer a un conjunto de evidencias recopiladas por los forenses, para identificar entre los sospechosos de una violación, al actor del crimen; pudiera tratarse de cepas de virus para definir el patrón en que se infectó una población de individuos; o solo sea la muestra de un bebe que aun esta en el interior de su madre para conocer su sexo, o si presenta alguna enfermedad congénita. Cualquiera que sea el caso, es necesario realizar un análisis que permita identificar patrones para agrupar las muestras por comportamientos semejantes en conglomerados o clusters. Cuando tenemos pocas muestras, (10-20), es fácil realizar el análisis sin necesidad de una aplicación como soporte, pero ¿Qué ocurrirá cuando se tengan cientos de muestras? Pues que se estará propenso a errores en un mayor por ciento.

### 1.2.3 Análisis de Clusters

El Análisis Cluster es un conjunto de técnicas utilizadas para clasificar los objetos o casos en grupos homogéneos llamados *conglomerados (clusters)* con respecto a algún criterio de selección predeterminado. Los objetos dentro de cada grupo (conglomerado), son similares entre sí (alta homogeneidad interna) y diferentes a los objetos de los otros conglomerados o clusters (alta heterogeneidad externa). Es decir, que si la clasificación hecha es óptima, los objetos dentro de cada clusters estarán cercanos unos de otros y los clusters diferentes estarán muy apartados. Por ello, es también conocido como análisis de clasificación o taxonomía numérica. [9]

El Análisis Clusters tiene como propósito esencial, agrupar aquellos objetos que reúnan idénticas características, es decir, se convierte así en una técnica de análisis exploratorio diseñada para revelar las agrupaciones naturales dentro de una colección de datos. [10]

Los dos tipos de procedimientos de agrupación son los jerárquicos y los no jerárquicos. El cluster jerárquico se caracteriza por el desarrollo de una jerarquía o estructura de árbol (dendograma). De este modo, los clusters están formados solamente por la unión de los grupos existentes, así cualquier miembro de un clusters puede trazar su relación en un irrompible sendero que comenzaría con una simple relación. [11]

Como el cluster agrupa objetos similares, se necesita una medida para evaluar las diferencias y similitudes entre objetos. La *Similitud* (similitud) es una medida de correspondencia o semejanza entre los objetos que van a ser agrupados. Lo más común es medir la equivalencia en términos de la distancia entre los pares de objetos. Así, los objetos con distancias reducidas entre ellos son más parecidos entre sí que aquellos con distancias mayores y se agruparán por lo tanto, dentro del mismo clusters. [12]

### 1.3 Estado del Arte

Las aplicaciones que trabajan con los marcadores moleculares son muy diversas y es de esperar que cada vez se les encuentren nuevos usos. Por ahora se vienen empleando en la diferenciación de individuos, discriminación entre clones, análisis filogenéticos y taxonómicos, mapeo de genomas, cuantificación de variabilidad génica e ínter-específica, mejoras genéticas, detección de infecciones, localización de resistencia a enfermedades, y dispersión de especies. Entre estas aplicaciones están:

### 1.3.1 GelCompar

El GelCompar trabaja en la comparación de geles que están introducidos en bases de datos y realizando el análisis de los patrones genéticos de cada cepa que coinciden. La comparación y normalización de los geles la realiza por análisis de los marcadores de peso molecular presentes en cada gel. Para calcular la semejanza entre los perfiles y las distancias genéticas entre los aislamientos se pueden aplicar varios coeficientes de correlación y para estudiar la relación genética de las cepas posee un árbol de similitud basado en diferentes algoritmos. La semejanza entre los aislamientos se obtiene de la comparación densitométrica de un análisis completo de cada perfil y no únicamente de la presencia o ausencia de bandas.

Luego de eliminar aislamientos repetidos de un mismo paciente, define las cepas que provienen de una misma población como aislamientos <agrupamientos> si ellos compartieron patrones idénticos o con una sola banda de diferencia (como se sabe las secuencias de inserción son elementos genéticos móviles, por lo que entre cepas de una misma cadena de transmisión [clusters] puede ocurrir un cambio.) Si sus patrones genéticos son totalmente diferentes entre sí entonces no se realiza agrupamiento. Son aplicaciones muy completas pero que poseen un precio muy elevado en el mercado de hasta 2000 y 10000 euros el paquete.

### 1.3.2 BioNumeric

El BioNumeric y el GelCompar son el mismo software pero GelCompar se limita al análisis de la huella digital del gel. Mientras el BioNumeric puede analizar caracteres filogenéticos y datos de secuencia, así como datos de gel.

El GelCompar es un subconjunto del BioNumeric limitado para tomar las huellas digitales. Este tiene dos módulos: la cuantificación básica y comparativa y el análisis de polimorfismo. El BioNumeric tiene su propio motor de base de datos, que expresamente es diseñado para el almacenaje de los tipos diversos de datos personales. La capacidad de la base de datos es muy alta, mientras el mantenimiento es fácil, que lo hace la opción perfecta para el análisis local de un solo usuario, maneja bases de datos relacionadas, que usan un motor de base de datos externo, como el Oráculo o el Servidor SQL.

El BioNumeric tiene un módulo poderoso para el trabajo con clusters así como una gama de algoritmos para trabajar estos métodos. Presenta una descripción comprensiva de todos los experimentos disponibles para una selección de entradas y permite al usuario mostrar y comparar

cualquier combinación de experimentos, las agrupaciones obtenidas pueden ser comparadas con el modelo o caracteres obtenidos de otros experimentos. Al igual que el GelCompar son software muy completos pero muy caros.

### 1.3.3 GelQuant

GelQuant es un software de análisis de gel fácil de usar y económico. GelQuant lee imágenes de gel en cualquier formato así las imágenes pueden haber sido generadas de cualquier cámara o sistema de escáner.

### 1.3.4 TotalLab

El TotalLab es un software que realiza el análisis de geles, es muy fácil de usar. Los instrumentos son incluidos para el análisis tanto de cintas como de manchas de una variedad de muestras. TotalLab lee imágenes de gel en cualquier formato, las imágenes pueden haber sido generadas de cualquier cámara o el sistema de escáner.

#### Actividades que realiza

Análisis de geles de Electroforesis.

Análisis de manchas de Electroforesis de 2-D.

Análisis de Punto y Manchas de Ranura. [13]

### 1.3.5 Mat-Gen

Es un software especializado hecho en Cuba que permite estimar las relaciones genéticas entre pares de genotipos empleando marcadores moleculares o caracteres codificados en sistema binario. Brinda la opción de utilizar indicadores ajustados para el sesgo por presencia de genofondo de similitud genética común. Los estimados obtenidos por diferentes indicadores pueden ser comparados para cada combinación de pares de genotipos mediante el error Standard estimado por métodos de remuestreo. Permite analizar hasta 80 genotipos y hasta 20 bandas por sistema. Fue hecho para el trabajo con plantas y esta actualmente ya en desuso. [14]

#### ¿Por qué desarrollar una aplicación nacional?

El desarrollo de una aplicación nacional facilitaría el trabajo, a centros que realicen análisis de cluster

sobre imágenes de caracterización genética. Se implementaría con las funcionalidades que precise el Instituto de Medicina Tropical “Pedro Kourí”, teniendo como material de apoyo los softwares antes mencionados y el conocimiento de los especialistas cubanos. Implicaría a su vez un ahorro del presupuesto, que de otra manera, se tendría que utilizar para la compra de las licencias de dichos softwares comerciales.

### **1.4 Metodologías que pueden utilizarse para el desarrollo del sistema.**

La metodología en el desarrollo de un software es el plano de apoyo, la guía que nos conducirá a realizar un software de calidad, es el conjunto de métodos que se deben realizar para realizar un software con calidad. Con la selección y utilización de la metodología utilizada tendremos clientes más complacidos y desarrolladores satisfechos, además servirá para que el desarrollo y confección del software no parezca complicado, riguroso y difícil de controlar.

En un proyecto de desarrollo de software la metodología define ¿Quién debe hacer Qué?, ¿Cuándo y cómo debe hacerlo? Una metodología es un proceso. No existe una metodología de software universal. Las características de cada proyecto (equipo de desarrollo, recursos, etc.) exigen que el proceso sea configurable

#### **1.4.1 RUP**

El Proceso Unificado es un proceso de desarrollo de software que contiene un conjunto de actividades necesarias para transformar los requisitos de un usuario en un sistema de software. Más que un simple proceso; es un marco de trabajo genérico que puede especializarse para una gran variedad de sistemas de software, para diferentes áreas de aplicación, tipos de organizaciones, niveles de actitud y tamaños de proyecto. Constituye la metodología estándar más utilizada para el análisis, implementación y documentación de sistemas orientados a objetos. RUP no es un sistema con pasos firmemente establecidos, sino un conjunto de metodologías adaptables al contexto y necesidades de cada organización. Está basado en componentes, lo cuál quiere decir que el sistema software en construcción está formado por componentes de software interconectados a través de interfaces bien definidas. La metodología RUP, llamada así por sus siglas en inglés Rational Unified Process, se divide en 4 fases para el desarrollo del software:

**Inicio:** El Objetivo en esta etapa es determinar la visión del proyecto.

**Elaboración:** En esta etapa el objetivo es determinar la arquitectura óptima.

**Construcción:** En esta etapa el objetivo es llevar a obtener la capacidad operacional inicial.

**Transmisión:** El objetivo es llegar a obtener del proyecto.

Se considera que RUP es una muy buena metodología, pues permite mayor productividad de un equipo y la realización de mejores prácticas de software a través de plantillas y herramientas que lo guían en todas las actividades de desarrollo crítico del software.

Utiliza el Lenguaje Unificado de Modelado (Unified Modeling Language, UML) para preparar todos los esquemas de un sistema software. Garantiza la elaboración de todas las fases de un producto de software orientado a objetos.

UML es un lenguaje gráfico para visualizar, especificar, construir y documentar un sistema de software. UML ofrece un estándar para describir un "plano" del sistema (modelo), incluyendo aspectos conceptuales tales como procesos de negocios y funciones del sistema, y aspectos concretos como expresiones de lenguajes de programación, esquemas de bases de datos y componentes de software reutilizables.

El punto importante para notar aquí es que UML es un "lenguaje" para especificar y no un método o un proceso. UML se usa para definir un sistema de software; para detallar los artefactos en el sistema; para documentar y construir es el lenguaje en el que está descrito el modelo. UML se puede usar en una gran variedad de formas para soportar una metodología de desarrollo de software (tal como el Proceso Unificado de Rational). [15]

*Características del Proceso Unificado:*

Los verdaderos aspectos definitorios del Proceso Unificado, y que lo convierten en único, se resumen en tres frases clave - dirigido por casos de uso, centrado en la arquitectura, e iterativo e incremental.

**Dirigido por los casos de uso:**

Teniendo en cuenta que la razón de ser de un sistema es brindar servicios a los usuarios, RUP define caso de uso como el conjunto de acciones que debe realizar un sistema para dar un resultado de valor a un determinado usuario y los utiliza tanto para especificar los requisitos funcionales del sistema, como para guiar todos los demás pasos de su desarrollo, dígame diseño, implementación y prueba.

**Estar centrado en la arquitectura:**

La arquitectura es una vista del diseño completo con las características más importantes, dejando a un

lado los detalles. Esta no solo incluye las necesidades de los usuarios e inversores, sino también otros aspectos técnicos como el hardware, sistema operativo, sistema de gestión de base de datos, protocolos de red; con los que debe coexistir el sistema. En otras palabras, la arquitectura representa la forma del sistema, la cual va madurando en su interacción con los casos de uso hasta llegar a un equilibrio entre funcionalidad y características técnicas.

### **Ser iterativo e incremental:**

La alta complejidad de los sistemas actuales hace que sea factible dividir el proceso de desarrollo en varios mini-proyectos. Cada uno de estos mini-proyectos se les denomina iteración y pueden o no representar un incremento en el grado de terminación del producto completo. En cada iteración los desarrolladores seleccionan un grupo de casos de uso, los cuales se diseñan, implementan y prueban. La planificación de iteraciones hace que se reduzcan los riesgos de los costes de un solo incremento, no sacar al mercado un producto en el tiempo previsto, mantener la motivación del equipo pues puede ver avances claros a corto plazo y que el desarrollo pueda adaptarse a los cambios en los requisitos. [16]

### **1.4.2 Metodología XP (Extreme Programming)**

Mientras que RUP intenta reducir la complejidad del software por medio de estructura y la preparación de las tareas pendientes en función de los objetivos de la fase y actividad actual, XP, como toda metodología ágil, lo intenta por medio de un trabajo orientado directamente al objetivo, basado en las relaciones interpersonales y la velocidad de reacción.

XP intenta minimizar el riesgo de fallo del proceso por medio de la disposición permanente de un representante competente del cliente a disposición del equipo de desarrollo. Este representante deberá estar en condiciones de contestar rápida y correctamente a cualquier pregunta del equipo de desarrollo de forma que no se retrase la toma de decisiones, de ahí lo de incompetente.

A partir de las de la arquitectura perseguida se crea un plan de liberación o entrega del software, entre el equipo de desarrollo y el cliente. Para cada entrega se discutirán los objetivos de la misma con el representante del cliente y se definirán las iteraciones (de pocas semanas de duración) necesarias para cumplir con los objetivos de la entrega. El resultado de cada iteración es un programa que se transmite al cliente para que lo juzgue. En base a su opinión se definen las siguientes iteraciones del proyecto y si el cliente no está contento se adaptará el plan de entregas e iteraciones hasta que el cliente de su aprobación y el software este a su gusto.

En XP se programará solo la funcionalidad que es requerida para la entrega actual. Es decir, una gran flexibilidad y capacidad de configuración solo será implementada cuando sea necesaria para cumplir los requerimientos de la entrega. Se sigue un diseño evolutivo con la siguiente premisa: conseguir la funcionalidad deseada de la forma más sencilla posible. Este diseño evolutivo hace que no se le de apenas importancia al análisis como fase independiente, puesto que se trabaja exclusivamente en función de las necesidades del momento. [17]

### **1.4.3 Metodología FDD (Desarrollo guiado por la funcionalidad, Feature driven development en ingles)**

FDD se podría considerar a medio camino entre RUP y XP, aunque al seguir siendo un proceso ligero es más similar a este último. Está pensado para proyectos con tiempo de desarrollo relativamente cortos (menos de un año). Se basa en un proceso iterativo con iteraciones cortas que producen un software funcional que el cliente y la dirección de la empresa pueden ver y monitorizar.

Un proyecto que sigue FDD se divide en 5 fases:

1. Desarrollo de un modelo general.
2. Construcción de la lista de funcionalidades.
3. Plan de entrega en base a las funcionalidades a implementar.
4. Diseñar en base a las funcionalidades.
5. Implementar en base a las funcionalidades.

Se hará uso de las herramientas de la metodología RUP (Rational Unified Process) para facilitar el desarrollo del sistema, porque, es muy pesada, dividido en cuatro fases, las fases se dividen en iteraciones, los artefactos son el objetivo de cada actividad y se basa en roles, utiliza UML, necesita de mucha documentación y es muy organizativa. [18]

### **1.5 Herramientas utilizadas para el desarrollo del sistema.**

**Herramientas CASE** (ingeniería de software asistida por computadora, en inglés, Computer Aided Software Engineering.)

**“Se puede definir a una herramienta CASE como un conjunto de programas y ayudas que dan**



**asistencia a los analistas, ingenieros de software y desarrolladores, durante todos los pasos del ciclo de vida de desarrollo de un software.” [19]**

Ventajas con la utilización de las herramientas CASE:

Permiten el incremento en la velocidad de desarrollo de los sistemas.

Permiten a los analistas tener más tiempo para el análisis y diseño y minimizar el tiempo para codificar y probar.

En las etapas del proceso de desarrollo de software permiten:

Automatizar el dibujo de diagramas.

Ayudar en la documentación del sistema.

Ayudar en la creación de relaciones en la base de datos.

Generar estructuras de código.

Aumentan la productividad. Esto se consigue a través de la automatización de determinadas tareas, como la generación de código y la reutilización de objetos o módulos. [20]

### **1.5.1 Rational Rose Enterprise Edition**

Es una herramienta para “modelado visual”, que forma parte de un conjunto más amplio de herramientas que juntas cubren todo el ciclo de vida del desarrollo de software. Permite completar una gran parte de las disciplinas (flujos fundamentales) del proceso unificado de Rational (RUP) e incluye un conjunto de herramientas de ingeniería inversa y generación de código que allanan el camino hasta el producto final. [21]

El Rational es una herramienta CASE basada en UML que permite crear los diagramas que se van generando durante el proceso de ingeniería en el desarrollo del software. Es completamente compatible con la metodología RUP, brinda muchas facilidades en la generación de la documentación del software que se está desarrollando, además posee un gran número de estereotipos predefinidos que facilitan el proceso de modelación del software. Es capaz de generar el código fuente de las clases definidas en el flujo de trabajo de diseño, pero tiene la limitación de que aún hay varios lenguajes de programación que no soporta o que sólo lo hace a medias.

Rational se encuentra a la cabeza en cuanto al desarrollo del Unified Modeling Language (UML), que se ha convertido en la notación estandarizada empleada en Rational Rose para especificar, visualizar y construir desarrollos de software y sistemas.

Rose es una herramienta con plataforma independiente que ayuda a la comunicación entre los miembros del equipo, a monitorear el tiempo de desarrollo y a entender el entorno de los sistemas. Una de las grandes ventajas de Rose es que permite a los arquitectos de software y desarrolladores visualizar el sistema completo utilizando un lenguaje común, además los diseñadores pueden modelar sus componentes e interfaces en forma individual y luego unirlos con otros componentes del proyecto. [Rational] [22]

### 1.6.2 Visual paradigma – UML

Visual Paradigm para UML es una de las herramientas UML CASE del mercado, considerada como muy completa y fácil de usar, con soporte multiplataforma y que proporciona excelentes facilidades de interoperabilidad con otras aplicaciones. Fue creada para el ciclo vital completo del desarrollo del software que lo automatiza y acelera, permitiendo la captura de requisitos, análisis, diseño e implementación. Visual Paradigm-uml también proporciona características tales como generación del código, ingeniería reversa y generación de informes. Tiene la capacidad de crear el esquema de clases a partir de una base de datos y crear la definición de base de datos a partir del esquema de clases. Permite invertir código fuente de programas, archivos ejecutables y binarios en modelos UML al instante, creando de manera simple toda la documentación. Está diseñada para usuarios interesados en sistemas de software de gran escala con el uso del acercamiento orientado a objeto, además apoya los estándares más recientes de las notaciones de Java y de UML. Incorpora el soporte para trabajo en equipo, que permite que varios desarrolladores trabajen a la vez en el mismo diagrama y vean en tiempo real los cambios hechos por sus compañeros.

En este trabajo se utilizará el Rational Rose, por las ventajas ya explicadas anteriormente. [23]

### 1.6 Roles de RUP desempeñados en el desarrollo de la aplicación:

Los roles son una definición abstracta de un grupo de responsabilidades que deben llevar a cabo ciertas actividades del proceso y producir algunos documentos. No son individuos, ni son necesariamente equivalentes a títulos profesionales. En su lugar, los roles describen cómo los individuos a los que se les asignan deberán comportarse en el contexto del proyecto.

En el Proceso Unificado de Desarrollo de Software los roles han sido clasificados en grupos, de acuerdo a las actividades que realizan.

**Analista:** Agrupa los roles que están involucrados fundamentalmente en la extracción e investigación de los requisitos del sistema. Este grupo está formado por los siguientes roles:

**Analista de procesos del negocio:** Responsable de definir la arquitectura del negocio; los casos de uso del negocio y actores, así como sus relaciones. Los artefactos que realizan son: modelos de casos de uso del negocio, modelo de análisis del negocio, reglas del negocio, glosario del negocio, documento de arquitectura del negocio, documento visión, metas del negocio, especificación suplementaria del negocio y objetivos del negocio.

**Diseñador de negocio:** Encargado de detallar la especificación de la organización o parte de ella. Los artefactos que realizan son: actor del negocio, caso de uso del negocio, realización de caso de uso del negocio, trabajador del negocio, entidad del negocio, sistema del negocio.

**Analista del sistema:** Dirige y coordina el proceso de extracción de requisitos y desarrollo del modelo de casos de uso, definiendo la funcionalidad y límites del sistema. Los artefactos que realizan son: plan de gestión de requerimientos, documento visión, modelo de casos de uso, glosario.

**Especificador de requerimientos:** Se encarga de especificar los detalles de una o varias partes de la funcionalidad del sistema, describiendo uno o varios aspectos de los requisitos. Los artefactos que realizan son: actor, casos de uso, paquete de casos de uso y especificación de requerimiento del software.

### Actividades

Una actividad en concreto es una unidad de trabajo que una persona que desempeñe un rol puede ser solicitado a que realice. Las actividades tienen un objetivo concreto, normalmente expresado en términos de crear o actualizar algún producto.

### Flujos de trabajo

Se le llama a una lista de actividades, trabajadores y artefactos que constituyen un proceso. Es una secuencia de actividades que produce un resultado valioso.

**Modelado del negocio:** Con este flujo de trabajo se pretende llegar a un mejor entendimiento de la organización donde se va a establecer el producto.

**Requerimientos:** Este es uno de los flujos de trabajo más importantes, porque en él se establece qué tiene que hacer exactamente el sistema que se construya. En esta línea los requisitos son el contrato que se debe cumplir, de modo que los usuarios finales tienen que comprender y aceptar los requisitos que se especifiquen.

**Análisis y Diseño:** El análisis consiste en obtener una visión del sistema que se preocupa de ver qué hace, de modo que sólo se interesa por los requisitos funcionales. Por otro lado el diseño es un refinamiento del análisis que tiene en cuenta los requisitos no funcionales, en definitiva cómo cumple el sistema sus objetivos.

En este trabajo se llegará hasta el flujo de trabajo análisis y diseño.

Se utilizarán varios Artefactos, entiéndase por ello la información producida, modificada o usado durante el proceso de desarrollo de software. Los productos son los resultados tangibles del proyecto, las cosas que va creando y usando hasta obtener el producto final.

**Reglas del negocio:** Describen políticas que deben cumplirse o condiciones que deben satisfacerse, por lo que regulan algún aspecto del negocio.

**Glosario del negocio:** Lista de concepto asociados al negocio que son comúnmente usados y que deben ser del dominio del equipo de desarrollo para poder modelar el negocio y dar una solución a la problemática encontrada.

**Modelo de casos de uso del sistema:** Es un modelo del sistema que contiene actores, casos de uso y sus relaciones.

**Glosario:** Términos comunes que se utilizan para describir el sistema.

**Especificación de requerimiento del software:** Es la captura de requerimientos de software para el sistema o una parte de este.

**Analista de Sistema**

**Plan de Gestión de Requerimientos:** describe los artefactos de los requisitos, tipos de requisitos, y sus cualidades respectivas, especificando la información que será recogida y mecanismos de control que se utilizará para medir, divulgar, y cambios que controlan a los requisitos del producto.

**Glosario:** se definen los términos comunes importantes usados al describir el sistema.

**Especificación Suplementaria:** se especifican que no se capturan fácilmente en artefactos del comportamiento de los requisitos tales como especificaciones del caso a utilizar.

**Solicitudes del Stakeholder:** contiene cualquier tipo de peticiones que un stakeholder (cliente, usuario final, persona de la comercialización, y así sucesivamente) pudo tener en el sistema que se convertirá. Puede también contener referencias a cualquier tipo de fuentes externas a las cuales el sistema deba conformarse.

**Documento Visión:** define la opinión de los Stakeholders del producto que será desarrollado, especificado en términos de necesidades y características cruciales para los stakeholders.

**Modelo de Casos de Uso:** permite que los desarrolladores de software y los clientes lleguen a un acuerdo sobre los requisitos, sobre las condiciones y posibilidades que debe cumplir el sistema. Este modelo contiene actores, casos de uso y sus relaciones.

**Diseñador:** Identifica y define las responsabilidades, funcionamientos, atributos, y relaciones de elementos en el plan. Asegura que el plan es consistente con la arquitectura del software, y se detalla a un punto dónde la aplicación puede proceder.

Actividades que desempeña este rol:

**Casos de uso del análisis:** permite identificar las responsabilidades, atributos y asociaciones de las clases.

**Diseño de caso de uso:** permite refinar los requisitos en los funcionamientos de subsistemas del plan y/o sus interfaces.

**Clases del diseño:** se ocupa de los requisitos no funcionales relacionados a la clase.

Artefactos generados por este rol:

**Modelo de caso de uso:** describe cómo un caso del uso particular se comprende dentro del modelo del diseño.

**Clases del análisis:** representan un modelo conceptual para el sistema que tiene responsabilidades y conducta.

**Clases del diseño:** es una descripción de un juego de objetos que comparten las mismas responsabilidades, relaciones, funcionamientos, atributos, y semántica. [24]

### **1.7 Conclusiones parciales:**

Las aplicaciones que trabajan con muestras de caracterización genética son muchas y muy desarrolladas, por tanto podemos contar con una base muy sólida para el desarrollo de la aplicación.

Existen varias metodologías para el desarrollo de software pero RUP reúne todas las características necesarias para el desarrollo de la aplicación. Se analiza las tendencias y tecnologías actuales, dando los conceptos necesarios para comprender el por qué de las tecnologías que serán utilizadas en el desarrollo de la aplicación. Se fundamenta la metodología a utilizar UML así como los dos roles que se desempeñan con los que se abarca el análisis y diseño de la aplicación.

# CAPÍTULO 2

## Características del Sistema

### 2.1 Introducción

En el presente capítulo se hace la descripción de la propuesta que trae este trabajo, para ello se describen los procesos del negocio que tiene que ver con el objeto de estudio, de acuerdo a esto se llega a la conclusión que debido a la poca estructuración de esos procesos, para poder entender el contexto en que se emplaza el sistema necesitamos definir conceptos que podemos agrupar en un Modelo de Dominio, para capturar correctamente los requisitos y poder construir un sistema correcto.

Además se enumeran los requisitos funcionales y no funcionales que debe tener el sistema que se propone, lo que nos permite hacer una concepción general del sistema, e identificar mediante un Diagrama de Caso de Uso, las relaciones de los actores que interactúan con el sistema, y las secuencias de acciones con las que interactúan.

### 2.2 Modelo del Dominio.

Teniendo en cuenta el bajo nivel de estructuración que tiene el negocio que estamos estudiando, con soluciones muy diversas y dispersas, aunque todas llevan el mismo propósito de realizar el análisis de conglomerados en las imágenes de las muestras de caracterización genética. Se decidió hacer modelo de dominio, ya que permite de manera visual mostrar al usuario los principales conceptos que se manejan en el dominio del sistema en desarrollo. Esto ayuda a los usuarios, a utilizar un vocabulario común para poder entender el contexto en que se emplaza el sistema. Para capturar correctamente los requisitos y poder construir un sistema correcto se necesita tener un firme conocimiento del funcionamiento del objeto de estudio. Este modelo va a contribuir posteriormente a identificar algunas clases que se utilizarán en el sistema. [25]

#### 2.2.1 Definición de las entidades y los conceptos principales

Primeramente vamos a identificar todos los conceptos que se utilizarán en el diagrama, mediante un glosario de términos, sobre los nombres:

Se le denominará **Investigador** a cualquier persona que trabaje con la aplicación, sin importar su categoría.

Se considera **Imagen** a todos las imágenes digitales que se introduzcan en la aplicación, a la cual se le realiza todo el análisis.

Se denominará **Patrón** a todos los patrones que se obtienen de la imagen para realizar el análisis de cluster.

Se llamará **Informe de análisis** al documento que guardará el árbol resultante de realizar el cluster.

Se denomina **Comparación de muestras** a todos los procesos que se le realizará a la imagen para lograr los clusters.

Se llamará **Asignar Patrones** a la acción de introducir el peso molecular de cada banda de la muestra patrón en la imagen que se este analizando en ese momento.

El modelo del dominio se describe mediante diagramas UML, específicamente con un diagrama de clases conceptuales significativas en el dominio del problema.



### 2.2.2 Diagrama del modelo de dominio.

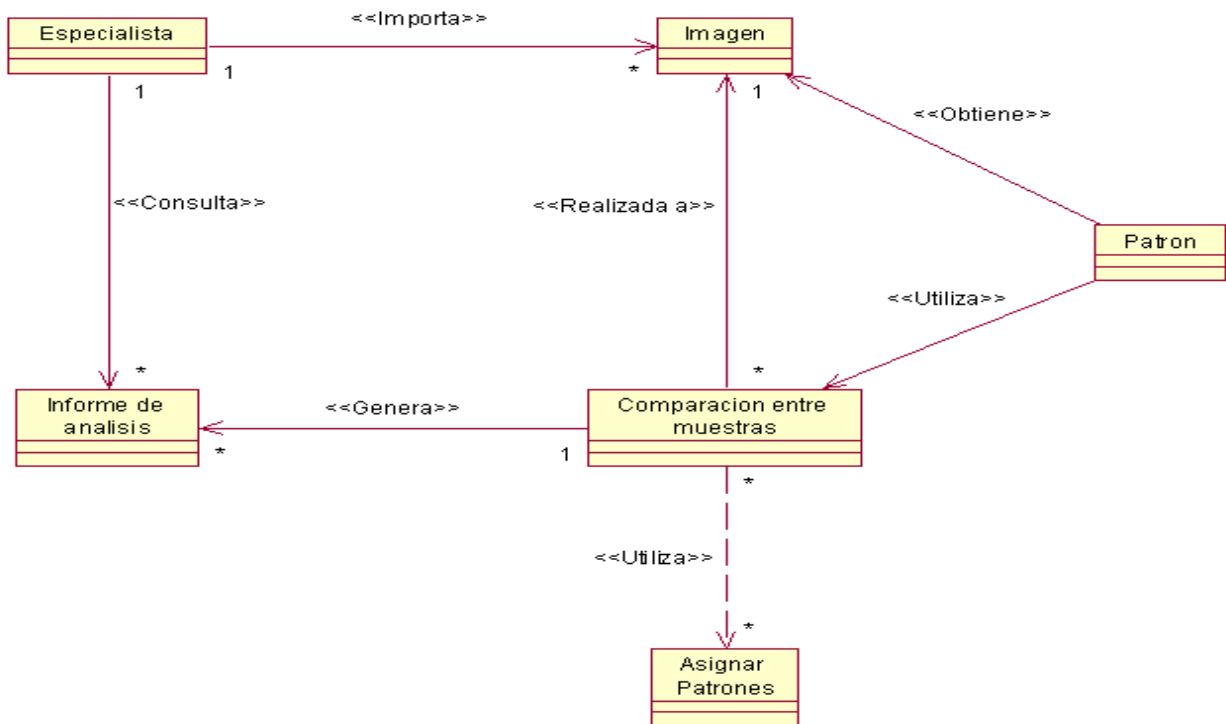


Fig. 1 Modelo de dominio

### 2.2.3 Reglas del negocio

El proceso de realizar clusters a imágenes de caracterización genética debe regirse por las siguientes reglas para su cumplimiento de manera exitosa. Ellas son:

Colocar como mínimo dos patrones por cada imagen y como máximo tres, uno en cada extremo y el tercero en el centro del gel.

Cada análisis debe tener como máximo 20 muestras de individuos.

El marcador molecular que se utilice, será a consecuencia, en dependencia del investigador.

El marcador molecular debe “barrer” la mayor cantidad de bandas que generen las muestras en el experimento.

## 2.3 Especificación de los Requisitos del Software

Los requerimientos del sistema es uno de los aspectos más importantes a considerar cuando se desarrolla un software, pues estos constituyen la base, el fundamento de la solución propuesta y el argumento para desarrollar el modelado del sistema. Por eso la captura de los requerimientos es uno de los procesos críticos de la ingeniería del software.

### 2.3.1 Requerimientos Funcionales

Una vez conocidos los conceptos que rodean al objeto de estudio, se puede empezar a analizar ¿Qué debe hacer el sistema para que se cumplan los objetivos planteados al inicio de este trabajo?, para ello enumeraremos a través de requerimientos funcionales las funciones que el sistema deberá ser capaz de realizar. Dentro de ellos se incluyen las acciones que podrán ser ejecutadas por el usuario, las acciones ocultas que debe realizar el sistema, y las condiciones extremas a determinar por el sistema. De acuerdo con los objetivos planteados el sistema debe ser capaz de:

**R.1** Preprocesamiento de imagen.

**1.1** Resaltado.

**1.2** Eliminación del fondo.

**1.3** Reconstrucción de la muestra.

**1.4** Separación y selección de la muestra.

**R.2** Asignar patrones.

**R.3** Normalizar Gel.

**R.4** Identificar Bandas.

**4.1** Eliminar Bandas.

**4.2** Insertar Bandas.

**4.3** Calcular peso molecular.

**R.5** Realizar análisis de cluster.

**5.1** Calcular similitud.

**5.2** Crear clusters.

**5.3** Crear dendograma

**R.6** Gestionar informe.

**6.1** Buscar informe.

**6.2** Eliminar informe.

**6.3** Guardar informe.

R.7 Gestionar gel.

7.1 Guardar gel

7.2 Eliminar gel.

7.3 Buscar gel.

7.4 Importar gel.

R.8 Gestionar muestras.

8.1 Insertar muestras.

8.2 Eliminar muestras.

8.3 Buscar muestras.

8.4 Guardar muestras.

### **2.3.2 Requerimientos No Funcionales.**

Los requerimientos no funcionales son propiedades o cualidades que el producto debe tener. Debe pensarse en estas propiedades como las características que hacen al producto atractivo, usable, rápido o confiable.

#### **1. Requerimientos de apariencia o interfaz externa.**

La aplicación estará diseñada de modo tal que el usuario tenga el mayor control posible de la misma, simple de usar y legible.

#### **2. Requerimientos de usabilidad.**

La aplicación estará orientada para que se use por especialistas que no poseen muchos conocimientos informáticos. Por lo que tendrá un buen nivel de uso una vez instalado en el IPK. Su uso permitirá agilizar el proceso de obtener cluster de imágenes de caracterización genética.

#### **3. Requerimientos de soporte**

El sistema debe propiciar su mejoramiento y la anexión de otras opciones que se le incorporen en un futuro. Así como un manual de usuario para el mejor entendimiento del usuario.

#### **4. Requerimientos de portabilidad.**

El sistema será multiplataforma.

#### **5. Requerimientos de confiabilidad.**

El sistema no debe presentar fallos y en casos de alguno debe garantizar que las pérdidas de información sean mínimas. Además no cuelga al sistema operativo.

### **6. Requerimientos de instalación**

Para la instalación de la aplicación se debe instalar una maquina virtual de Java y tener dos instaladores, uno para la aplicación y otro para la Base de datos.

### **7. Requerimientos de software.**

Se debe disponer de cualquier versión de Windows98 o superior y SQL Server 2005 para la implantación de la BD.

### **8. Requerimientos de hardware.**

Para el desarrollo y puesta en práctica del proyecto se requieren máquinas con los siguientes requisitos:

256MB RAM

Procesador PENTIUM

40 GB HDD

Se debe contar con una impresora para imprimir el informe generado del análisis de cluster.

## 2.4 Diagrama de Casos de Uso del Sistema

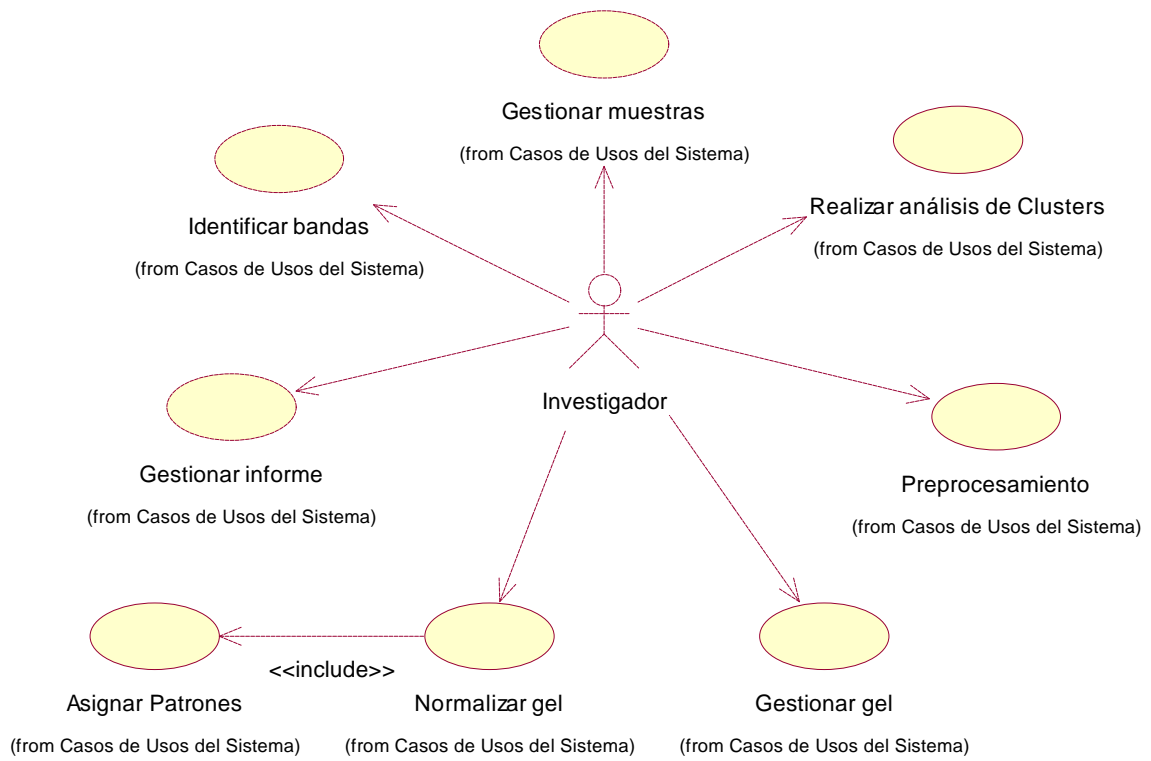


Fig. 2 Diagrama Caso de Uso del Sistema.

### 2.4.1 Descripción de los casos de uso.

Los casos de uso se utilizan para obtener información de cómo debe trabajar el sistema, son descripciones de la funcionalidad del sistema independiente de la implementación, describen bajo la forma de acciones y reacciones el comportamiento de un sistema desde el punto de vista del usuario. [Anexo 1]

Los casos de usos definidos son los siguientes:

- ✍ Preprocesamiento.
- ✍ Asignar patrones.
- ✍ Normalizar gel.
- ✍ Identificar Bandas

- ✍ Realizar análisis de cluster.
- ✍ Gestionar informe.
- ✍ Gestionar gel.
- ✍ Gestionar muestras.

**2.4.2 Definición de los actores.**

Actores	Justificación
Investigador	Es el que inicia todos los casos de uso.

**2.4.3 Listado de casos de uso.**

<b>CU-1</b>	Preprocesamiento
<b>Actor</b>	Investigador
<b>Descripción</b>	Se realiza el preprocesamiento a la imagen, y de manera sucesiva se va realizando una serie de actividades sobre la imagen que permiten un mejor trabajo sobre la misma.
<b>Referencia</b>	R.1, R 1.1, R 1.2, R 1.3, R 1.4

<b>CU-2</b>	Asignar Patrones
<b>Actor</b>	Investigador
<b>Descripción</b>	Se selecciona(n) la(s) muestra(s) de referencia para introducirle los pesos moleculares a cada banda identificada por el sistema
<b>Referencia</b>	R.2

<b>CU-3</b>	Normalizar gel
-------------	----------------

<b>Actor</b>	Investigador
<b>Descripción</b>	Se ajusta la posición de las bandas respecto al sistema de referencia.
<b>Referencia</b>	R.3

<b>CU-4</b>	Identificar Bandas
<b>Actor</b>	Investigador
<b>Descripción</b>	Se realiza la identificación de las bandas y se calcula el peso molecular de cada muestra.
<b>Referencia</b>	R.4, R 4.1, R 4.2, R 4.3

<b>CU-5</b>	Realizar análisis de Clusters.
<b>Actor</b>	Investigador
<b>Descripción</b>	Se realiza el agrupamiento de las muestras analizadas.
<b>Referencia</b>	R.5, R 5.1, R 5.2, R 5.3

<b>CU-6</b>	Gestionar informe
<b>Actor</b>	Investigador
<b>Descripción</b>	Se guarda, busca, o elimina informes de la base de datos.
<b>Referencia</b>	R.6, R 6.1, R 6.2, R 6.3

<b>CU-7</b>	Gestionar gel
-------------	---------------

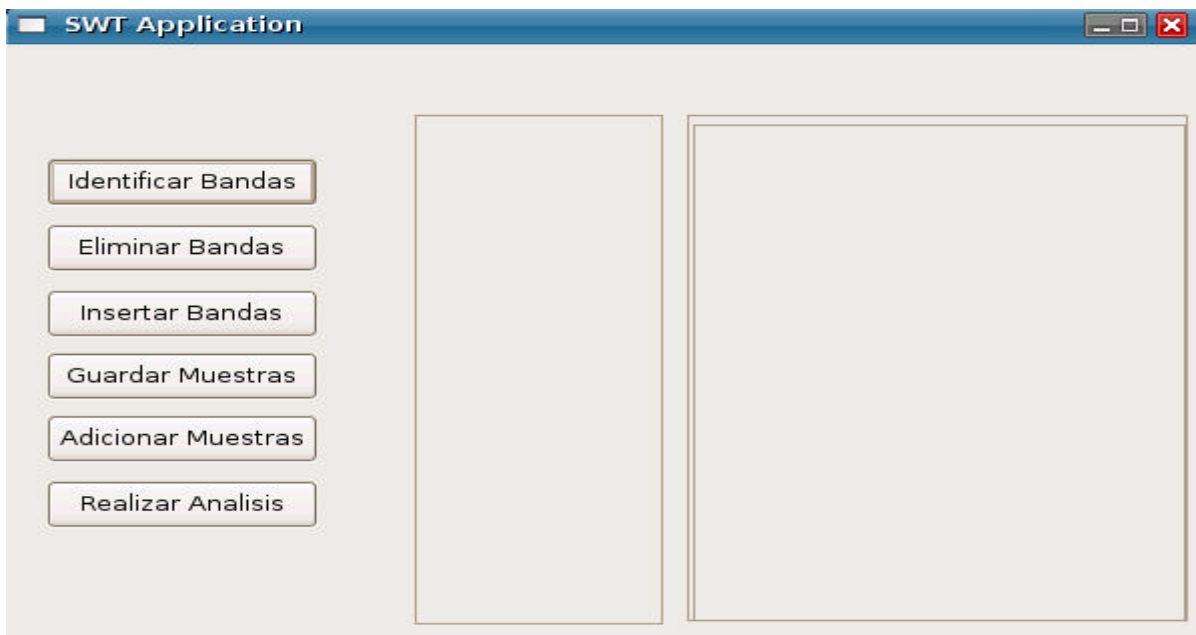
<b>Actor</b>	Investigador
<b>Descripción</b>	Se guarda, busca, o elimina el gel en la base de datos.
<b>Referencia</b>	R.7, R 7.1, R 7.2, R 7.3, R 7.4

<b>CU-8</b>	Gestionar muestras
<b>Actor</b>	Investigador
<b>Descripción</b>	Se inserta, busca, guarda o elimina las muestras en la base de datos.
<b>Referencia</b>	R.8, R 8.1, R 8.2, R 8.3, R 8.4

## 2.5 Prototipos de interfaz de usuario.







### **2.6 Conclusiones parciales**

En este capítulo se realizó el modelo de dominio al sistema, ya que el negocio no está bien definido. Se analizaron los requerimientos funcionales y no funcionales del sistema, respetando las solicitudes y necesidades de los clientes. Por último se diseñó el modelo del sistema con 9 casos de uso, donde se plantearon todas las funcionalidades que el sistema debe cumplir y se hizo una descripción textual de todos los casos de usos especificando cuales eran los críticos. Así como se presento los prototipos de interfaz de la aplicación.

# CAPÍTULO 3

## Análisis y Diseño del Sistema

### 3.1 Introducción.

En este capítulo se desarrolla el flujo de análisis y diseño, es uno de los flujos de trabajo del RUP, metodología empleada en la modelación del sistema. Aunque esta metodología une al Análisis y Diseño en un solo flujo de trabajo ambos tienen funcionalidades diferentes. Estos dos procesos son muy importantes en el desarrollo de software dirigido por modelos ya que constituyen la vista lógica de la arquitectura del software.

Se realiza el análisis del sistema, modelando su principal diagrama (Diagrama de Análisis) y el diagrama de clases del diseño que es uno de los artefactos principales del flujo de trabajo del RUP. Además se presenta el modelo de datos que es la base para construir la base de datos que soportará el trabajo del sistema. Finalmente después de modelar la lógica del negocio a través de las clases, se tratan los principios del diseño de la aplicación, donde se desarrolla el Diagrama de Clases de Diseño, el Modelo de Clases Persistente, el Modelo de Datos y el Modelo de Despliegue.

### 3.2 Análisis

El análisis consiste en obtener una visión del sistema que se preocupa por ver ¿Qué hace?, de modo que sólo se interesa por los requisitos funcionales, transformar los requisitos funcionales en un diseño de clases viendo las relaciones e interacción que existe entre ellos, y teniendo en cuenta en el proceso una arquitectura robusta que permita adaptar el sistema al entorno de implementación que se está desarrollando.

#### 3.2.1 Modelo conceptual de clases de análisis

El modelo conceptual de clases del análisis tiene las siguientes características:

Descrito con el lenguaje del desarrollador.

Estructurado por clases.

Utilizado por los desarrolladores para comprender como debería ser diseñado el sistema.

Esboza como llevar acabo la funcionalidad dentro del sistema, sirve como una primera aproximación al diseño.

Define realizaciones de casos de uso, y cada una de ellas representa el análisis de un caso de uso del modelo de casos de uso.

### 3.3 Diagrama de Clases de Análisis

Una de las principales salidas o artefactos del análisis es el Diagrama de Clases de Análisis. En este se representan las Clases de Análisis (clase Interfaz, clase controlador y clases entidad), sus relaciones entre sí y con los actores del sistema. [Anexo 2]

### 3.4 Principios de diseño

#### Interfaz al usuario

La interfaz diseñada para el sistema está basada en el estándar de ventanas. El tipo de letra a utilizar será *Arial* de estilo regular y tamaño 8 y el diseño de la aplicación deber ser adecuado para el tipo de usuario (Investigador), logrando la fácil comprensión del mismo en el lenguaje utilizado para las opciones que brindan. El sistema debe mostrar una barra de menú en la parte derecha donde deben estar la mayoría de las opciones de trabajo.

#### Ayuda

Una parte importante de cualquier sistema lo constituye la ayuda, en el caso nuestro en la barra de menú aparecerá una opción de ayuda con temas relacionado a la utilización de la aplicación con el objetivo de aclarar alguna duda al especialista. En la misma debe aparecer información con imágenes para facilitar la comprensión, también debe presentar la posibilidad de acceder a la misma con solo presionar la tecla F1.

#### Tratamiento de errores

En el diseño de la interfaz se debe tener en cuenta el tratamiento de errores logrando que los mensajes de error que emita el sistema sean de fácil comprensión para el usuario y lo más

descriptivos posibles y además debe alertarlos de posibles riesgos de las operaciones que realice.

### 3.5 Patrones

Un patrón es una pareja de **problema / solución** con un nombre, que codifica (estandariza) buenos principios y sugerencias. Es aplicable a otros contextos, con una sugerencia sobre la manera de usarlo en diferentes situaciones. El objetivo de los patrones es crear un lenguaje común a una comunidad de desarrolladores para comunicar experiencia sobre los problemas y sus soluciones. Pueden referirse a distintos niveles de abstracción, desde un proceso de desarrollo hasta la utilización eficiente de un lenguaje de programación. [26]

#### 3.5.1 Patrones de arquitectura

Para el diseño de la arquitectura se emplea el patrón de 3 capas porque:

- Se navega de las capas superiores a las inferiores.
- Se pueden hacer modificaciones en una capa sin afectar a otras.
- La capa de presentación es la página que el usuario puede ver.
- La capa de negocio es la formada por las clases controladoras, las de acceso a datos y las clases entidades.
- La capa de datos es la formada por la Base de Datos. [27]

#### 3.5.2 Patrones de diseño

Para el diseño se utilizan los patrones de asignación de responsabilidades (GRASP, Patrones de Software para la asignación General de Responsabilidad, en english Acrónimo de General Responsibility Assignment Software Patterns.):

- ? Alta cohesión: Reduce la dependencia, la información que almacena una clase debe de ser coherente y está en la mayor medida de lo posible relacionada con la clase. La responsabilidad se reparte en varias clases y existe una fuerte colaboración entre ellas, es muy eficiente. [28]
- ? Bajo acoplamiento: Las clases más independientes, reduce el impacto al cambio y más reutilizables.

- ? Controlador: Evita el acceso directo a las clases entidades, evita que la capa de presentación no maneje los eventos del sistema. [29]

### 3.6 Diseño

Con el diseño se logra la comprensión de los aspectos relacionados con los requisitos no funcionales y restricciones relacionadas con los lenguajes de programación, componentes reutilizables, Sistemas Operativos, tecnologías de distribución y concurrencia y tecnologías de interfaz de usuario y se crea una entrada apropiada y un punto de partida para actividades de implementación, capturando los requisitos o subsistemas individuales, interfaces y clases.

Uno de las principales salidas del diseño lo constituyen el Diagrama de Clases del Diseño, el Diagrama de Clases Persistentes, el Modelo de Datos y el Diagrama de Despliegue.

#### 3.6.1 Diagrama de Clases del Diseño

Un Diagrama de Clases muestra un conjunto de clases, interfaces y colaboraciones, así como sus relaciones.

Los diagramas de clases se utilizan para modelar la vista de diseño estática de un sistema, esto incluye modelar el vocabulario del sistema, modelar las colaboraciones o modelar esquemas. Los diagramas de clases también son la base para un par de diagramas relacionados: los diagramas de componentes y los diagramas de despliegue. Los diagramas de clases son importantes no sólo para visualizar, especificar y documentar modelos estructurales, sino también para construir sistemas ejecutables, aplicando ingeniería directa e inversa. [30] [Anexo 3]

#### 3.6.2 Descripción de las clases.

<b>Nombre:</b> Interfaz de trabajo	
<b>Tipo de clase:</b> interfaz	
<b>Atributo</b>	<b>Tipo</b>
1. Importar_gel	Button
2. Sistema _ referencia	Button
3. Separacion_muestras	Button
4. Guardar_gel	Button
5. Buscar_gel	Button

6. Buscar _ informe	Button
7. Preprocesamiento	Button
<b>Para cada responsabilidad:</b>	
Nombre:	Descripción:
1. Mostrar_gel	Es la operación que muestra el gel para ser analizada.
2. Importar_gel	Es la operación que importa un gel de un directorio de la pc a la aplicación.

<b>Nombre:</b> Normalizacion	
<b>Tipo de clase:</b> interfaz	
<b>Atributo</b>	<b>Tipo</b>
1. Introducir_peso_molecular	Button
2. Selección_patron_referencia	Button
3. Seleccionar_muestras_referencias	Button
4. Asignar_posicion_referencia	Button
5. Indentificar_bandas	Button
6. Normalizar	Button
7. Guardar_peso_molecular	Button
<b>Para cada responsabilidad:</b>	
Nombre:	Descripción:

<b>Nombre:</b> Identificar bandas	
<b>Tipo de clase:</b> interfaz	
<b>Atributo</b>	<b>Tipo</b>
1. Indentificar_bandas	Button
2. Insertar _ bandas	Button
3. Eliminar _ bandas	Button
4. Realizar_analisis	Button
5. Guardar _ muestras	Button
6. Adicionar _ muestras	Button
<b>Para cada responsabilidad:</b>	

Nombre:	Descripción:
1. Mostrar_mensaje_error	Es la operación que muestra un mensaje de error en caso entrar mal algún dato.
2. Mostrar_muestras	Es la operación que muestra las muestras.

<b>Nombre:</b> Analisis	
<b>Tipo de clase:</b> interfaz	
<b>Atributo</b>	<b>Tipo</b>
1. Generar _ informe	Button
2. Realizar_analisis	Button
3. Imprimir	Button
<b>Para cada responsabilidad:</b>	
Nombre:	Descripción:
1. Mostrar_analisis	Es la operación que muestra el análisis después de realizado.
2. Mostrar_mensaje_error	Es la operación que muestra el mensaje de error en caso de que se entre un dato mal.

<b>Nombre:</b> Buscar muestras	
<b>Tipo de clase:</b> interfaz	
<b>Atributo</b>	<b>Tipo</b>
1. Nombre _ muestras	TextField
2. Id _ muestras	TextField
3. Provincia/Estado	TextField
4. Localidad	TextField
5. Tipo _ muestras	TextField
6. Pais origen	TextField
7. Buscar _ muestras	Button
8. Eliminar _ muestras	Button
<b>Para cada responsabilidad:</b>	
Nombre:	Descripción:
1. Mostrar_mensaje_error	Es la operación que muestra el mensaje de error en caso de que se halla entrado algún dato mal.



2. Mostrar_muestras	Es la operación que muestra las muestras en la interfaz
3. Insertar_muestras	Es la operación que envía un mensaje a la controladora para que inserte las muestras.
4. Buscar_muestras	Es la operación que envía un mensaje a la controladora para que busque las muestras.
5. Eliminar_muestras	Es la operación que envía un mensaje a la controladora para que elimine las muestras.

<b>Nombre:</b> Buscar informe	
<b>Tipo de clase:</b> interfaz	
<b>Atributo</b>	<b>Tipo</b>
1. Nombre _ investigador	TextField
2. Id _ informe	TextField
3. Fecha _ análisis	TextField
4. Buscar _ informe	TextField
5. Eliminar _ informe	TextField
<b>Para cada responsabilidad:</b>	
Nombre:	Descripción:
1. Mostrar_informe	Es la operación que muestra el informe después de buscado.
2. Mostrar_mensaje_error	Es la operación que muestra el mensaje de error en caso de que se entre algún dato con error.

<b>Nombre:</b> Buscar gel	
<b>Tipo de clase:</b> interfaz	
<b>Atributo</b>	<b>Tipo</b>
1. Nombre_gel	TextField
2. Id_gel	TextField
3. Analizar	Button
4. Buscar_gel	Button
5. Eliminar_gel	Button
<b>Para cada responsabilidad:</b>	
Nombre:	Descripción:

1. Mostar_mensaje_error	Muestra un mensaje de error si se entra mal algún dato.
2. Mostrar_gel	Es la operación que muestra el gel después de buscarlo en la base de datos.
3. Buscar_gel	Es la operación que envía un mensaje a la controladora para que busque el gel.
4. Eliminar_gel	Es la operación que envía un mensaje a la controladora para que elimine el gel.

<b>Nombre:</b> Gestionar analisis	
<b>Tipo de clase:</b> Control	
<b>Atributo</b>	<b>Tipo</b>
<b>Para cada responsabilidad:</b>	
<b>Nombre:</b>	<b>Descripción:</b>
1 .Imprimir	Es la operación que permite que se imprima el informe.
2.Calcular_similitud	Es la operación que permite calcular la similitud entre las muestras.
3. Realizar_analisis	Es la operación que permite llamar el objeto Analisis_cluster donde se implementan todos los métodos que lleva este proceso.
4. Crear_clusters	Es la operación que genera los clusters entre las muestras.
5. Crear_dendograma	Es el método que permite crear el dendograma.
6. Buscar_peso_molecular	Es la operación que realiza la búsqueda de los pesos moleculares de cada muestra

<b>Nombre:</b> Gestionar informe	
<b>Tipo de clase:</b> Control	
<b>Atributo</b>	<b>Tipo</b>
<b>Para cada responsabilidad:</b>	
<b>Nombre:</b>	<b>Descripción:</b>
1.Buscar_informe	Es la operación que busca el informe en la base de datos.
2.Eliminar_informe	Es la operación que permite eliminar el informe de la base de datos.
3. Guardar_informe	Es la operación que guarda el informe en la base de datos.

4. Enviar_mensaje_error	Es la operación que lanza un mensaje si existe un error en la entrada de datos.
-------------------------	---

<b>Nombre:</b> Gestionar_gel	
<b>Tipo de clase:</b> Control	
<b>Atributo</b>	<b>Tipo</b>
<b>Para cada responsabilidad:</b>	
<b>Nombre:</b>	<b>Descripción:</b>
1.Guardar_gel	Es la operación que permite guardar un nuevo gel a la base de datos.
2.Buscar_gel	Es la operación que permite buscar geles en la base de datos.
3.Eliminar_gel	Es la operación que permite eliminar geles de la base de datos.
4. Importar_gel	Es la operación que permite cargar el gel desde un directorio cualquiera de la pc.
5. Enviar_mensaje_error	Es la operación que lanza un mensaje si existe un error en la entrada de datos.

<b>Nombre:</b> Gestionar_muestras	
<b>Tipo de clase:</b> Control	
<b>Atributo</b>	<b>Tipo</b>
<b>Para cada responsabilidad:</b>	
<b>Nombre:</b>	<b>Descripción:</b>
1.Insertar _ muestras	Es la operación que permite adicionar nuevas muestras al análisis.
2.Buscar _ muestras	Es la operación que permite buscar muestra en la base de datos.
3.Eliminar _ muestras	Es la operación que permite eliminar muestra de la base de datos.
4. Guardar_muestras	Es la operación que permite guardar una nueva muestra en la base de datos.
5. Enviar_mensaje_error	Es la operación que lanza que hay un error por una mala entrada de datos.
6. Buscar_peso_molecular	Es la operación que busca el peso molecular de las muestras.

<b>Nombre:</b> Acceso a datos	
<b>Tipo de clase:</b> Control	
<b>Atributo</b>	<b>Tipo</b>
1.Nombre_Base_Datos	String
2. Id_conexion	Integer
3. Nombre_Host	String
4. Consulta	String
5. Mensaje_error	String
<b>Para cada responsabilidad:</b>	
<b>Nombre:</b>	<b>Descripción:</b>
1.Mostrar_error	Es la operación que permite mostrar mensajes cuando existe un dato con error.
2.Buscar_muestra	Es la operación que permite realizar la consulta de buscar muestra en la base de datos.
3.Eliminar_muestra	Es la operación que permite realizar la consulta de eliminar muestra de la base de datos.
4. Buscar_gel	Es la operación que permite realizar la consulta de buscar geles en la base de datos.
5. Adicionar_gel	Es la operación que permite realizar la consulta de adicionar geles en la base de datos.
6. Eliminar_gel	Es la operación que permite realizar la consulta de eliminar geles de la base de datos.
7. Buscar_informe	Es la operación que permite realizar la consulta de buscar informe en la base de datos.
8. Eliminar_informe	Es la operación que permite realizar la consulta de eliminar informe en la base de datos.
9. Abrir_conexion	Es la operación que abre la conexión a la base de datos
10. Cerrar_conexion	Es la operación que cierra la conexión a la base de datos
11. Guardar_patron	Es la operación que permite realizar la consulta de guardar el peso molecular del patrón en la base de datos.
12. Buscar_patron	Es la operación que permite realizar la consulta de buscar el peso molecular del patrón en la base de datos.
13. Guardar_PM_muestras	Es la operación que permite realizar la consulta de guardar el peso

	molecular de las muestras en la base de datos.
14. Buscar_peso_molecular	Es la operación que permite realizar la consulta de buscar el peso molecular de las muestras en la base de datos.
15. Guardar_muestra	Es la operación que permite realizar la consulta de guardar muestra en la base de datos.
16. Guardar_informe	Es la operación que permite realizar la consulta de guardar informe en la base de datos.

<b>Nombre:</b> Procesamiento de imagen	
<b>Tipo de clase:</b> Control	
<b>Atributo</b>	<b>Tipo</b>
<b>Para cada responsabilidad:</b>	
<b>Nombre:</b>	<b>Descripción:</b>
1. Resaltado	Es la operación que permite resaltar el contraste del gel.
2. Eliminar_fondo	Es la operación que permite eliminar el fondo del gel.
3. Reconstruccion	Es la operación que permite reconstruir la muestra.
4. Selección_separacion	Es la operación que selecciona y separa las muestras.
5. Preprocesamiento	Es la operación que llama a los demás métodos.

<b>Nombre:</b> Gestionar bandas	
<b>Tipo de clase:</b> Control	
<b>Atributo</b>	<b>Tipo</b>
<b>Para cada responsabilidad:</b>	
<b>Nombre:</b>	<b>Descripción:</b>
1. Identificar_bandas	Es la operación que permite identificar las bandas para realizar el análisis.
2. Eliminar_bandas	Es la operación que permite eliminar bandas del gel.
3. Insertar_bandas	Es la operación que permite insertar bandas en el gel.

4. Buscar_patron	Es la operación que permite buscar el patrón del gel.
5. Calcular_PM_muestras	Es la operación que permite calcular los pesos moleculares de las muestras.
6. Guardar_PM_muestras	Es la operación que permite guardar los pesos moleculares calculados

<b>Nombre:</b> Normalizar	
<b>Tipo de clase:</b> Control	
<b>Atributo</b>	<b>Tipo</b>
<b>Para cada responsabilidad:</b>	
<b>Nombre:</b>	<b>Descripción:</b>
1. Introducir_peso_molucular	Es el método que permite introducir el peso molecular a las bandas patrones.
2. Seleccionar_patron_referencia	Es el método que permite seleccionar el patrón de referencia del gel.
3. Seleccionar_muestra_referencia	Es el método que permite seleccionar la muestra de referencia del gel.
4. Asignar_posicion_referencia	Es el método que permite asignar la posición de referencia de cada banda de las muestras patrones.
5. Identificar_bandas	Es el método que permite direccionar a la interfaz Identificar_bandas.
6. Normalizar	Es el método que permite la normalización de cada gel.
7. Guardar_patron	Es el método que permite guardar el peso molecular del patrón.

<b>Nombre:</b> Patrón	
<b>Tipo de clase:</b> Entidad	
<b>Atributo</b>	<b>Tipo</b>
1. Peso molecular	Integer
2. Id_patron	Integer
<b>Para cada responsabilidad:</b>	
<b>Nombre:</b>	<b>Descripción:</b>

### 3.6.2 Diagramas de Interacción

Los Diagramas de Interacción se utilizan para modelar los aspectos dinámicos de un sistema. La mayoría de las veces, esto implica modelar instancias concretas o prototípicas de clases, interfaces, componentes y nodos, junto con los mensajes enviados entre ellos, todo en el contexto de un escenario que ilustra un comportamiento. Los Diagramas de Interacción pueden utilizarse para visualizar, especificar, construir y documentar la dinámica de una sociedad particular de objetos, o se pueden utilizar para modelar un flujo de control particular de un Caso de Uso. [31]

Permite el modelado dinámico del sistema. Representa “cómo” se comunican los objetos y se limita a un determinado contexto: Escenario, Subsistema. Estos diagramas son de dos tipos Diagrama de Secuencia o Colaboración.

#### 3.6.2.1 Diagrama de Colaboración

El Diagrama de Colaboración tiene como objetivo describir el comportamiento dinámico del sistema. Hace énfasis en la interacción entre objetos y como están vinculados. Muestra la misma información que un diagrama de secuencia pero de diferente modo. Aporta una visión estructural del Sistema. Este tipo de diagrama muestra las interacciones entre objetos organizadas entorno a los objetos y los enlaces entre ellos. El Diagrama de Colaboración ofrece una mejor visión del escenario cuando el analista está intentando comprender la participación de un objeto en el sistema.

Proporcionan la representación principal de un escenario, ya que las colaboraciones se organizan entorno a los enlaces de unos objetos con otros. En un diagrama de colaboración existen en consecuencia tres tipos de elementos:

- Objetos
- Vínculos
- Mensajes [32]

Los Diagramas de Colaboración se desarrollan para cada Caso de Uso

### 3.6.2.2 Diagrama de Secuencia

Un Diagrama de Secuencia es un Diagrama de Interacción que representa las interacciones entre objetos ordenadas en secuencia temporal. Los clientes entienden fácilmente este tipo de diagramas. Contiene detalles de implementación del escenario, incluyendo los objetos y clases que se usan para implementar el escenario, y mensajes pasados entre los objetos. En particular muestra los objetos participantes en la interacción y la secuencia de mensajes intercambiados. [33] [Anexo 2]

### 3.7 Diseño de la Base de Datos.

El diseño de la BD está compuesto fundamentalmente por dos artefactos: El Modelo de Clases Persistentes y el Modelo de Datos. En este caso el modelo de datos tiene tres tablas: imagen, muestras e informe. A continuación se realiza la descripción de cada una. [Anexo 5]

Nombre de la tabla: Imagen		
Descripción general de qué datos almacena la tabla: Los atributos que posee permiten identificar, nombrar y guardar la imagen.		
Atributo	Tipo	Descripción
1. Id_gel	Entero	Es la llave de la tabla, con la que se identifica la imagen almacenada.
2. Nombre_gel	String	Es el nombre de cada una de las imágenes.
3. Imagen	String	Es el atributo que guarda el camino donde se encuentra la imagen.
4. Imagen normalizada.	String	Es el atributo que guarda el camino donde se encuentra la imagen ya normalizada.

Nombre de la tabla: Informe		
Descripción general de qué datos almacena la tabla: Los atributos que posee permiten identificar el informe, la fecha en que fue creado, por quien fue hecho y el numero de muestras que intervinieron en el experimento.		
Atributo	Tipo	Descripción
1. Id_informe	Entero	Es la llave de la tabla, con la que se identifica el informe almacenado.
2. Nombre_investigador	String	Es el nombre del investigador que realiza el análisis de cluster en las muestras.



3. Fecha	Date	Es el atributo que especifica el día/mes/año que se realizó el experimento.
4. Numero_muestras	Entero	Es el atributo que especifica el número de muestras presentes en el experimento.

Nombre de la tabla: Muestras		
<b>Descripción general de qué datos almacena la tabla:</b> Los atributos que posee permiten identificar el informe, la fecha en que fue creado, por quien fue hecho y el numero de muestras que intervinieron en el experimento.		
Atributo	Tipo	Descripción
1. Id_muestras	Entero	Es la llave de la tabla, con la que se identifica cada muestra.
2. Nombre_muestra	String	Es el nombre de la muestra.
3. Localidad	String	Es el atributo que especifica la localidad de donde proviene la muestra.
4. Pais_origen	String	Es el atributo que especifica el país de origen de la muestra.
5. Provincia/Estado	String	Son atributos con los que se especifica el país o el estado de la muestra.
6. Tipo_muestra	String	Es el atributo que especifica que tipo de muestra se usó en el análisis.
7. Peso_molecular	Integer	Es el atributo que especifica el peso de la muestra.
8. Fecha_aislamiento	Integer	Es el atributo que especifica la fecha en que fue aislada la muestra.

### 3.8 Modelo de Despliegue

Un Diagrama de Despliegue es un diagrama que muestra la configuración de los nodos que participan en la ejecución. Se modelan la topología del hardware sobre el que se ejecuta el sistema y la distribución física del sistema.

Para la puesta en práctica de la aplicación se necesitará principalmente un procesador que servirá

como servidor de Base de Datos conectado a las PCs clientes donde correrá la aplicación mediante el protocolo HTTP.

Las PCs clientes utilizarán una impresora siendo esta un recurso compartido conectada a la red mediante el puerto USB. [Anexo 6]

### **3.9 Generalidades de la implementación**

- Todos los campos de textos deben de ser validados en la interfaz antes de ser enviados a la BD.
- Se debe implementar un mecanismo de seguridad para bloquear la BD en caso de ataques.
- Todos los métodos deben nombrarse con mayúscula al inicio.

### **3.10 Conclusiones parciales.**

En este capítulo se representaron las diferentes partes de la solución propuesta a través de diferentes diagramas para modelar los elementos relativos a la aplicación.

Han sido definidas las clases de análisis y las clases del diseño así como sus relaciones en el Diagrama de Clases del Diseño. Se definió el Diagrama de Clases Persistentes, y a través de este, el Modelo de Datos. Además se confeccionó el Diagrama de Despliegue.

Se plantearon los principios de diseño de la aplicación, la concepción del Tratamiento de Errores y la concepción general de la Ayuda del Sistema.

## Conclusiones

Para llevar a cabo el modelamiento de la aplicación se realizó el análisis de algunas soluciones existentes en el mundo con características similares al sistema propuesto.

El sistema se diseñó siguiendo la metodología RUP y se utilizaron representaciones para la modelación de todas las fases del proyecto excepto para las de: Implementación y Prueba, las cuales no constituyen objetivos del trabajo.

En el desarrollo del trabajo se realizó el modelamiento del dominio y del sistema propuesto delimitado por el alcance del trabajo. Seguidamente se realizó el análisis de la aplicación la cual facilitó la comprensión y modelamiento del diseño de la propuesta. El sistema resultante esta provisto de un ambiente cómodo, fácil de entender, que cumple los estándares del diseño.

Se puede decir que se han cumplido exitosamente los objetivos trazados, modelándose una aplicación a la altura de lo que se quería y se necesitaba. Además, se cumplió estrictamente con los requisitos solicitados por el cliente.

## Recomendaciones.

Con el objetivo de continuar y mejorar la aplicación propuesta se recomienda:

- ✍ Realizar la implementación de la aplicación, para la solución real de los problemas existente en el Centro de Medicina Tropical Pedro Kourí.
- ✍ Continuar el estudio del tema con el objetivo de encontrar nuevas funcionalidades para futuras versiones de la aplicación.
- ✍ Permitir que otros centros que utilicen técnicas de caracterización genética puedan acceder a la aplicación.

## Bibliografía

- [1]. *La huella digital genética cumple 20 años.* [En línea] 2004. <http://www.consumer.es/web/es/educacion/2004/09/13/108774.php>.
- [2]. **Gutiérrez, Lic. Jaime Robleta.** Ciencias Penales . *El A.D.N y su importancia en la investigación criminalística.* . [En línea] 2000. [Citado el: 9 de 11 de 2006.]
- [3]. **Venter JC, Adams MD, Myers EW.** The sequence of the human genome.Science. 2001.
- [4]. **ANNIA FERRER MEDINA1, MARCELO NAZABAL GALVEZ1, LIDIA INES NOVOA PEREZ1.** Una aproximación a la genómica.
- [5]. Método STR para la Prueba de Paternidad por ADN. [En línea] [Citado el: 11 de 12 de 2006.] <http://biogenomica.com/STR.htm>.
- [6]. **Baginsky, Cecilia.** *Uso de Marcadores Moleculares, aplicaciones y ejemplos.* [En línea] [http://www.argenbio.org/h/biblioteca/libro/09\\_II\\_4.pdf](http://www.argenbio.org/h/biblioteca/libro/09_II_4.pdf).
- [7]. **Diaz, M. Gonzalo Claros.** *Marcadores moleculares:Qué son,cómo se obtiene y para qué valen.* [En línea] <http://www.encuentros.uma.es/encuentros49/marcadores.html>.
- [8]. *Los marcadores genéticos en la mejora vegetal.* Valencia: Universidad Politécnica de Valencia. **Carrillo, F. Nuez y J.M.** 2000.
- [9]. **Picca, Aurora, y otros.** *Marcadores Moleculares.* [En línea] [http://www.argenbio.org/h/biblioteca/libro/09\\_II\\_4.pdf](http://www.argenbio.org/h/biblioteca/libro/09_II_4.pdf).
- [10]. **Gutiérrez, Lic. Jaime Robleta.** Ciencias Penales . *El A.D.N y su importancia en la investigación criminalística.* . [En línea] 2000. [Citado el: 9 de 11 de 2006.] <http://www.poder-judicial.go.cr/salatercera/revista/REVISTA%2018/robleto18.htm>.
- [11]. Cluster Anayisis. [En línea] [http://www.datum.com.pe/Datum\\_english/cluster.pdf](http://www.datum.com.pe/Datum_english/cluster.pdf).
- [12]. **Nores, José Emilio Gondar.** *Análisis Cluster.* [En línea] 2000. <http://www.estadistico.com/arts.html?20001023> citada.
- [13]. *BioSystematica Gel Documentation & Gel Analysis.* [En línea] [Citado el: 8 de 2 de 2007.] <http://www.biosystematica.com/bionumerics.htm>.

- [14]. **Maria Teresa Cornide, Orlando Coto Arbelo.** *Manual de usuario del MAT-GEN versión 2.0* . 1998.
- [15]. Metodologías de desarrollo. [En línea] [http://www.informatizate.net/articulos/metodologias\\_de\\_desarrollo\\_de\\_software\\_07062004.html](http://www.informatizate.net/articulos/metodologias_de_desarrollo_de_software_07062004.html).
- [16]. *Ingeniería de Software*. [En línea] [Citado el: 22 de 3 de 2007.] <http://www.angelfire.com/scifi/jzavalar/apuntes/IngSoftware.html#ProcUnificado>.
- [17]. *Ingeniería de Software*. [En línea] [Citado el: 22 de 3 de 2007.] <http://www.angelfire.com/scifi/jzavalar/apuntes/IngSoftware.html#ProcUnificado>.
- [18]. **Jose H. Canós, Patricio Letelier y Maria Carmen Penadés.** *Metodologías Ágiles en el desarrollo de Software*. [En línea] <http://www.willydev.net/descargas/prev/TodoAgil.Pdf>.
- [19]. **Molpeceres, Alberto.** *Proceso de desarrollo:RUP,XP y FDD*. [En línea] 2002. [Citado el: 22 de 3 de 2007.] [http://www.javahispano.org/download/articulos/metodos\\_desarrollo.pdf;jsessionid=327D9C9F1F719F1F8FFFC0321C992B3E](http://www.javahispano.org/download/articulos/metodos_desarrollo.pdf;jsessionid=327D9C9F1F719F1F8FFFC0321C992B3E).
- [20]. *Herramientas CASE*. [En línea] [Citado el: 9 de 3 de 2007.] <http://www.cyta.com.ar/biblioteca/bddoc/bdlibros/proyectoinformatico/libro/c5/c5.htm>.
- [21]. *Rational Rose Enterprise Edition, la herramienta optima para UML*. [En línea] [Citado el: 9 de 3 de 2007.] [http://www.ciao.es/Rational\\_Rose\\_Enterprise\\_Edition\\_\\_Opinion\\_612900](http://www.ciao.es/Rational_Rose_Enterprise_Edition__Opinion_612900).
- [22]. **Jose H. Canós, Patricio Letelier y Maria Carmen Penadés.** *Metodologías Ágiles en el desarrollo de Software*. [En línea] <http://www.willydev.net/descargas/prev/TodoAgil.Pdf>.
- [23]. **Mendoza, María A.** *Metodología de desarrollo de software*. [En línea] Junio de 2004. [http://www.informatizate.net/articulos/metodologias\\_de\\_desarrollo\\_de\\_software\\_07062004.html](http://www.informatizate.net/articulos/metodologias_de_desarrollo_de_software_07062004.html).
- [24]. *10 Reasons to choose Visual Paradigm*. [En línea] [Citado el: 9 de 3 de 2007.] <http://www.visual-paradigm.com/aboutus/10reasons.jsp> .
- [25]. *Rational Enterprise Edition, Ayuda extendida*.
- [26]. *UML en 24 horas*.
- [27]. *Estilos y Patrones en la Estrategia de Arquitectura de Microsoft*. [En línea]

[http://www.microsoft.com/spanish/msdn/arquitectura/roadmap\\_arq/style.asp](http://www.microsoft.com/spanish/msdn/arquitectura/roadmap_arq/style.asp).

[28]. *Conferencia 1. Introducción a la Ingeniería de Software*. **Autores., Colectivo de**. Universidad de las Ciencias Informáticas : s.n., 2005-2006.

[29]. **Larman, Craig**. *UML y Patrones*.

[30]. **Facultad de informática, Universidad de Madrid**. *Patrones del Gang of Four*.

[31]. **Jacobson, I. y Booch, G. y Rumbaugh, J.** *El Proceso Unificado de Desarrollo de Software, Volumen I y II*. La Habana : s.n., 2004.

[32]. *Diagramas de Interacción según Métrica V3*. [En línea]  
<http://www.di.uniovi.es/~tuya/is/descarga/desarrollo/complem/DiagramasInteraccion.pdf>.

[33]. *Consultoría en Seguridad, diagramas de colaboración*. [En línea]  
<http://www.creangel.com/uml/interaccion.php>.

[34]. *Consultoría en seguridad*. [En línea] <http://www.creangel.com/uml/secuencia.php>.



## Anexos

### Anexo # 1: Descripción de los Casos de Uso del Sistema.

**Tabla 1 Descripción del Caso de uso Preprocesamiento.**

<b>Caso de uso:</b>	Preprocesamiento
<b>Actores:</b>	Investigador (inicia).
<b>Propósito:</b> Realizar operaciones sobre la imagen que permita un mejor trabajo con la misma.	
<b>Descripción:</b> Se realiza el preprocesamiento a la imagen, y de manera sucesiva se va realizando una serie de actividades sobre la imagen que permiten un mejor trabajo sobre la misma	
<b>Referencias:</b>	R.1, R 1.1, R 1.2, R 1.3, R 1.4
<b>Precondiciones:</b>	El gel debe estar cargado en la aplicación
<b>Postcondiciones:</b>	
<b>Requerimientos especiales:</b>	
<b>Curso Normal de los eventos</b>	
<b>Acciones del Actor</b>	<b>Respuesta del sistema</b>
1. El caso de uso se inicia cuando el Investigador pulsa el botón "Preprocesamiento".	1.1 Se resalta la imagen iniciándose el CU Resaltado. 1.2 Se elimina el fondo de la imagen iniciándose el CU Eliminar Fondo. 1.3 Se reconstruye la muestra iniciándose el CU Reconstrucción de muestra. 1.4 Se separa y selecciona la muestra iniciándose el CU Separación/Selección finalizando el CU.

Tabla 2 Descripción del Caso de uso Resaltado.

<b>Caso de uso:</b>	Resaltado
<b>Actores:</b>	
<b>Propósito:</b> Resaltar el gel.	
<b>Descripción:</b> Se resalta el contraste del gel para definir mejor las bandas.	
<b>Referencias:</b>	R 1.1
<b>Precondiciones:</b>	El caso de uso Preprocesamiento debe estar iniciado
<b>Postcondiciones:</b>	
<b>Requerimientos especiales:</b>	
<b>Curso Normal de los eventos</b>	
<b>Acciones del Actor</b>	<b>Respuesta del sistema</b>
	1. Se resalta la imagen finalizando el CU.

Tabla 3 Descripción del Caso de uso Eliminar fondo.

<b>Caso de uso:</b>	Eliminar fondo
<b>Actores:</b>	
<b>Propósito:</b> Eliminar el fondo del gel para su mejor trabajo.	
<b>Descripción:</b> El sistema después de resaltar el gel, elimina el fondo de la imagen.	
<b>Referencias:</b>	R 1.2
<b>Precondiciones:</b>	El gel debe estar resaltado
<b>Postcondiciones:</b>	
<b>Requerimientos especiales:</b>	
<b>Curso Normal de los eventos</b>	
<b>Acciones del Actor</b>	<b>Respuesta del sistema</b>
	1. El sistema elimina el fondo del gel finalizando el CU.

Tabla 4 Descripción del Caso de uso Reconstrucción de muestras.

<b>Caso de uso:</b>	Reconstrucción de muestra
<b>Actores:</b>	
<b>Propósito:</b> Reconstruir las muestras.	
<b>Descripción:</b> El sistema después de eliminar el fondo, reconstruye la muestra.	
<b>Referencias:</b>	R 1.3
<b>Precondiciones:</b>	El gel debe tener eliminado el fondo
<b>Postcondiciones:</b>	
<b>Requerimientos especiales:</b>	
<b>Curso Normal de los eventos</b>	
<b>Acciones del Actor</b>	<b>Respuesta del sistema</b>
	1. El sistema reconstruye las muestras finalizando el CU.

Tabla 5 Descripción del Caso de uso Separación/Selección.

<b>Caso de uso:</b>	Separación/Selección.
<b>Actores:</b>	
<b>Propósito:</b> Identificar las muestras y separarlas.	
<b>Descripción:</b> El sistema después de reconstruir las muestras, selecciona y separa las muestras.	
<b>Referencias:</b>	R 1.4
<b>Precondiciones:</b>	Las muestras deben estar reconstruida
<b>Postcondiciones:</b>	
<b>Requerimientos especiales:</b>	
<b>Curso Normal de los eventos</b>	
<b>Acciones del Actor</b>	<b>Respuesta del sistema</b>
	1. Define la ubicación de cada muestra.
	2. Separa las muestras finalizando el CU.

Tabla 6 Descripción del Caso de uso Normalizar Imagen.

<b>Caso de uso:</b>	Normalizar gel
<b>Actores:</b>	Investigador (inicia).
<b>Propósito:</b> Permite ajustar la posición de referencia de las bandas respecto al sistema de referencia.	
<b>Descripción:</b> Se ajusta la posición de las bandas respecto al sistema de referencia.	
<b>Referencias:</b>	R.2, R.3
<b>Precondiciones:</b>	Deben estar asignados los pesos moleculares a las bandas patrones.
<b>Postcondiciones:</b>	Las bandas patrones quedan ajustadas.
<b>Requerimientos especiales:</b>	
<b>Curso Normal de los eventos</b>	
<b>Acciones del Actor</b>	<b>Respuesta del sistema</b>
1. El caso de uso comienza cuando el Investigador oprime el botón "Normalizar".	1.1 Se inicia el CU asignar patrones. 1.3 Se ajusta la posición de las bandas patrones respecto al sistema de referencia, finalizando el CU.(Normalizar)
<b>Curso alternativo de los eventos:</b>	
<b>Prioridad:</b>	Crítico

Tabla 7 Descripción del Caso de uso Asignar patrones.

<b>Caso de uso:</b>	Asignar patrones.
<b>Actores:</b>	Investigador (inicia).
<b>Propósito:</b> Introducir los pesos moleculares a las bandas patrones.	
<b>Descripción:</b> Se selecciona(n) la(s) muestra(s) de referencia para introducirle los pesos moleculares a las bandas que el sistema identifica en la muestra seleccionada.	
<b>Referencias:</b>	R.2
<b>Precondiciones:</b>	La imagen debe estar preprocesada.
<b>Postcondiciones:</b>	Se obtiene el sistema de referencia
<b>Requerimientos especiales:</b>	
<b>Curso Normal de los eventos</b>	

Acciones del Actor	Respuesta del sistema
2. El Investigador selecciona la muestra que se tomara como patrón de de referencia.	2.1 Se muestra la opción de introducir los pesos moleculares de cada banda a la muestra seleccionada.
3. El Investigador introduce los pesos moleculares.	3.1 El sistema guarda los pesos moleculares del patrón, finalizando el CU.
<b>Curso alterno de los eventos:</b>	
<b>Prioridad:</b>	Secundario.

**Tabla 8 Descripción del Caso de uso: Identificar bandas.**

<b>Caso de uso:</b>	Identificar bandas.
<b>Actores:</b>	Investigador (inicia).
<b>Propósito:</b> calcula el peso molecular de cada muestra.	
<b>Descripción:</b> Identifican las bandas, y calcula los pesos moleculares a cada una de las muestras.	
<b>Referencias:</b>	R.4, R 4.1, R 4.2, R 4.3
<b>Precondiciones:</b>	El gel debe estar normalizado.
<b>Postcondiciones:</b>	
<b>Requerimientos especiales:</b>	
<b>Curso Normal de los eventos</b>	
Acciones del Actor	Respuesta del sistema
1. El Investigador pulsa el botón "Identificar bandas".	1.1 Identifica las bandas de cada muestra. 1.2 Busca el peso molecular del patrón asociado al gel. 1.3 Calcula el peso molecular de cada muestra. 1.4 Guarda el peso molecular de cada muestra finalizando el caso de uso.
<b>Curso alternativo de los eventos:</b>	

1.1 El investigador selecciona la banda que desea eliminar.	1.1.1 Elimina la banda
1.1 El investigador selecciona donde insertar la banda.	1.1.1 Inserta la banda
<b>Prioridad:</b>	Secundario.

**Tabla 9 Descripción del Caso de uso Realizar análisis de cluster.**

<b>Caso de uso:</b>	Realizar análisis de cluster.
<b>Actores:</b>	Investigador (inicia).
<b>Propósito:</b> Agrupar las muestras que sean semejantes entre ellas.	
<b>Descripción:</b> Se realiza el agrupamiento de las muestras presentes en el gel según el comportamiento semejante de cada una de ellas.	
<b>Referencias:</b>	R.5, R 5.1, R 5.2, R 5.3
<b>Precondiciones:</b>	Las muestras deben tener sus pesos moleculares.
<b>Postcondiciones:</b>	
<b>Requerimientos especiales:</b>	
<b>Curso Normal de los eventos</b>	
<b>Acciones del Actor</b>	<b>Respuesta del sistema</b>
1. El caso de uso se inicia cuando el Investigador pulsa el botón "Realizar análisis".	1.1 Busca el peso molecular de cada muestra. 1.2 Calcula la similitud entre las muestras. 1.2 Crea los clusters. 1.3 Crea el árbol resultante (dendograma). 1.4 Muestra el informe del análisis, finalizando el CU.
<b>Curso alternativo de los eventos:</b>	
1.1 El usuario pulsa el botón "imprimir"	1.2 Se imprime el informe de análisis.
<b>Prioridad:</b>	Crítico.

Tabla 10 Descripción del Caso de uso Realizar análisis de cluster.

<b>Caso de uso:</b>	Gestionar informe
<b>Actores:</b>	Investigador (inicia).
Propósito: Gestionar el proceso de actualización del los informes en la base de datos.	
<b>Descripción:</b> El caso de uso se inicia cuando se elije cualquiera de las opciones de guardar, buscar, o eliminar informe en la base de datos.	
<b>Referencias:</b>	R.6, R 6.1, R 6.2, R 6.3
<b>Precondiciones:</b>	
<b>Postcondiciones:</b>	Base de datos actualizada
<b>Requerimientos especiales:</b>	
<b>Curso Normal de los eventos</b>	
<b>Acciones del Actor</b>	<b>Respuesta del sistema</b>
1. El caso de uso se inicia cuando el Investigador decide " Gestionar informe."	1.1 Si el investigador desea guardar informe ver sección: guardar informe. Si el investigador desea buscar informe ver sección: buscar informe. Si el investigador desea eliminar informe ver sección: eliminar informe.
<b>Sección: guardar informe</b>	
1. El Investigador pulsa el botón "Guardar informe".	1.1 Muestra un formulario para guardar informe: Id_informe, nombre del investigador.
2. El investigador introduce los datos.	2.1 Verifica que el informe no exista en la base de datos. 2.2 Guarda el informe en la base de datos finalizando el CU.
<b>Sección: buscar informe</b>	
3. El investigador pulsa el botón "Buscar informe".	3.1 Muestra un formulario para buscar informe: id_informe, Nombre_investigador.

4. El investigador introduce los parámetros.	4.1 Verifica que exista(n) el(los) informe(es) en la base de datos. 4.2 Muestra un listado de informes.
5. El investigador selecciona el informe deseado.	5.1 Muestra el informe finalizando el CU.
<b>Sección: eliminar informe</b>	
6. El investigador pulsa el botón "Buscar informe".	6.1 Muestra un formulario para buscar informe: id_informe, Nombre_investigador.
7. El investigador introduce los datos.	7.1 Verifica que exista(n) el(los) informe(es) en la base de datos. 7.2 Muestra el listado de informes.
8. El investigador selecciona el informe a eliminar.	8.1 Se elimina(n) el(los) informe(s) seleccionada(s) de la base de datos finalizando el CU.
<b>Curso Alternativo de los eventos:</b>	
Acción 2.1 Si el informe existe en la base de datos se muestra un mensaje de error.	
Acción 4.1 Si el informe no existe en la base de datos se muestra un mensaje de error.	
Acción 7.1 Si el informe no existe en la base de datos se muestra un mensaje de error.	
5.1 El investigador pulsa el botón "imprimir".	5.1.1 Se imprime el informe.
<b>Prioridad:</b>	Secundario.

**Tabla 11 Descripción del Caso de uso Gestionar Imagen.**

<b>Caso de uso:</b>	Gestionar gel
<b>Actores:</b>	Investigador (inicia).
Propósito: Gestionar el proceso de actualización de un gel.	
<b>Descripción:</b> El caso de uso se inicia cuando se elige cualquiera de las opciones de guardar, buscar, o eliminar el gel de la base de datos.	
<b>Referencias:</b>	R.7, R 7.1, R 7.2, R 7.3,R 7.4
<b>Precondiciones:</b>	
<b>Postcondiciones:</b>	Base de datos actualizada
<b>Requerimientos especiales:</b>	
<b>Curso Normal de los eventos</b>	



Acciones del Actor	Respuesta del sistema
1. El Investigador desea "Gestionar gel".	1.1 Si el investigador desea insertar gel ver sección: guardar gel. Si el investigador desea buscar gel ver sección: buscar gel. Si el investigador desea eliminar gel ver sección: eliminar gel.
<b>Sección: Guardar gel</b>	
2. El investigador pulsa el botón "Guardar gel".	2.1 Muestra un formulario para guardar informe: Nombre_gel.
2. El investigador introduce los datos.	2.1 Verifica que el gel no exista en la base de datos. 2.2 Guarda el gel en la base de datos finalizando el CU.
<b>Sección: buscar gel</b>	
3. El investigador pulsa el botón "Buscar informe".	3.1 Muestra un formulario para buscar el gel: Id_gel, Nombre_gel.
4. El investigador introduce los parámetros.	4.1 Verifica que exista(n) el(los) gel(es) en la base de datos. 5.1 Muestra un listado de la búsqueda.
6. El investigador selecciona el gel deseado.	6.1 Muestra el gel finalizando el CU.
<b>Sección: eliminar gel</b>	
7. El investigador pulsa el botón "Buscar gel".	7.1 Muestra un formulario para buscar informe: id_informe, Nombre_investigador.
8. El investigador introduce los datos.	8.1 Verifica que exista(n) el(los) gel(es) en la base de datos. 8.2 Muestra el listado de búsqueda.
9. El investigador selecciona el(los) gel(es) a eliminar.	9.1 Se elimina(n) el(los) gel(es) seleccionado(s) de la base de datos finalizando el CU.
<b>Curso Alternativo de los eventos:</b>	

Acción 2.1 Si el gel existe en la base de datos se muestra un mensaje de error.	
Acción 4.1 Si el gel no existe en la base de datos se muestra un mensaje de error.	
Acción 8.1 Si el gel existe en la base de datos se muestra un mensaje de error.	
<b>Prioridad:</b>	Secundario.

**Tabla 12 Descripción del Caso de uso Gestionar Muestras.**

<b>Caso de uso:</b>	Gestionar muestra
<b>Actores:</b>	Investigador (inicia).
Propósito: Gestionar el proceso de actualización de las muestras	
<b>Descripción:</b> El caso de uso se inicia cuando se elige cualquiera de las opciones de adicionar, buscar, eliminar o cargar la(s) muestra(s) en la base de datos.	
<b>Referencias:</b>	R.8, R 8.1, R 8.2, R 8.3, R 8.4
<b>Precondiciones:</b>	
<b>Postcondiciones:</b>	Base de datos actualizada
<b>Requerimientos especiales:</b>	
<b>Curso Normal de los eventos</b>	
<b>Acciones del Actor</b>	<b>Respuesta del sistema</b>
1. El caso de uso se inicia cuando el Investigador desea "Gestionar muestra".	1.1 Si el investigador desea insertar muestra ver sección: insertar muestra. Si el investigador desea buscar muestra ver sección: buscar muestra. Si el investigador desea eliminar muestra ver sección: eliminar muestra.
<b>Sección: insertar muestra</b>	
2. El investigador pulsa el botón "Insertar muestras".	2.1 Muestra el formulario para buscar muestras: ID de la muestra, ubicación geográfica de origen, tipo de muestra, fecha de aislamiento.

3. El investigador introduce los datos.	3.1 Verifica que exista(n) la(s) muestra(s) en la base de datos. 3.2 Muestra el listado de muestras.
4. El investigador selecciona las muestras a insertar.	4.1 Se insertan las muestras finalizando el CU.
<b>Sección: buscar muestra</b>	
5. El investigador pulsa el botón "Buscar muestra".	5.1 Muestra un formulario para buscar muestras: ID de la muestra, ubicación geográfica de origen, tipo de muestra, fecha de aislamiento.
6. El investigador introduce los datos.	6.1 Verifica que exista(n) la(s) muestra(s) en la base de datos. 6.2 Muestra el listado de muestras finalizando el CU.
<b>Sección: eliminar muestra</b>	
7. El investigador pulsa el botón "Buscar muestra".	7.1 Muestra un formulario para buscar muestras: ID de la muestra, ubicación geográfica de origen, tipo de muestra, fecha de aislamiento.
8. El investigador introduce los datos.	8.1 Verifica que exista(n) la(s) muestra(s) en la base de datos. 8.2 Muestra el listado de muestras.
9. El investigador selecciona las muestras a eliminar.	9.1 Se elimina(n) la(s) muestra(s) seleccionada(s) finalizando el CU.
<b>Sección: Guardar muestras</b>	
10. El investigador pulsa el botón "Guardar muestras".	10.1 Muestra un formulario para guardar muestras: ID de la muestra, ubicación geográfica de origen, tipo de muestra, fecha de aislamiento.
11. El investigador introduce los datos.	11.1 Verifica que no exista la muestra en la base de datos. 11.2 Se guardan las muestras en la base de datos.

---

<b>Curso Alternativo de los eventos:</b>	
Acción 3.1 La muestra buscada no existen, se muestra un mensaje de error.	
Acción 6.1 La muestra buscada no existen, se muestra un mensaje de error.	
Acción 8.1 La muestra buscada no existen, se muestra un mensaje de error.	
Acción 11.1 La muestra buscada ya existe, se muestra un mensaje de error.	
<b>Prioridad:</b>	Secundario.

Anexo 2: Diagramas de clases del Análisis.

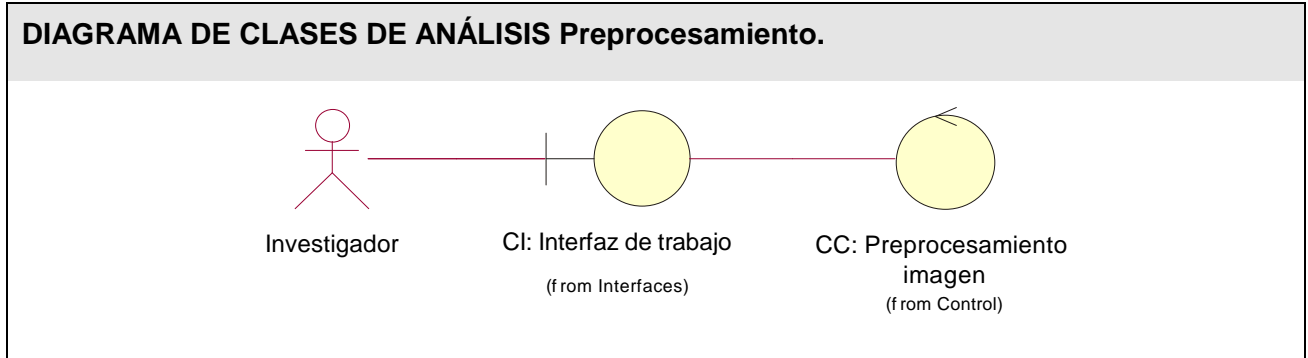


Fig. 1 Diagrama de análisis # 1: Preprocesamiento.

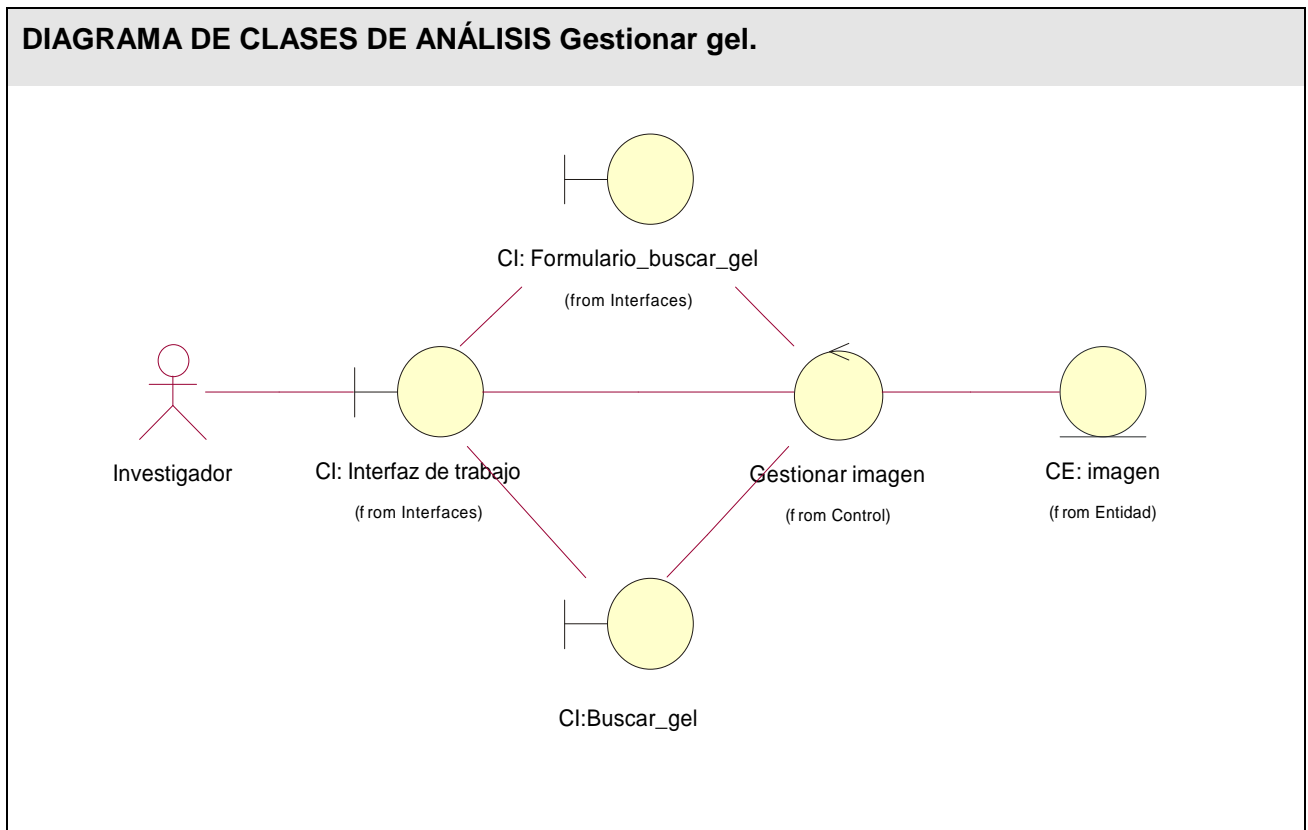


Fig. 2 Diagrama de análisis # 2: Gestionar gel.

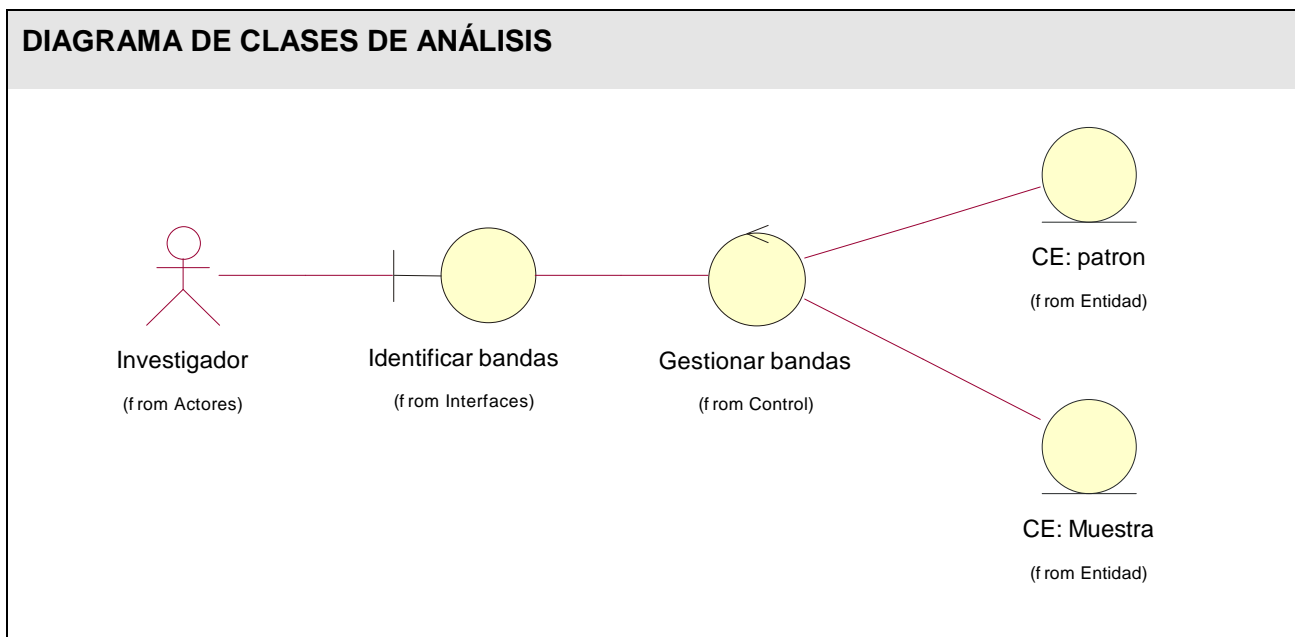


Fig. 3 Diagrama de análisis # 3: Identificar bandas patrones.

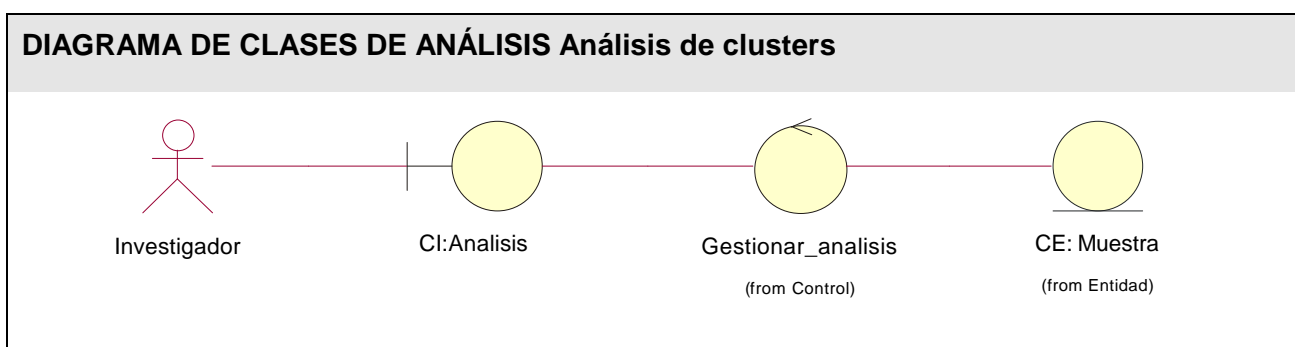


Fig. 4 Diagrama de análisis # 4: Análisis de cluster.

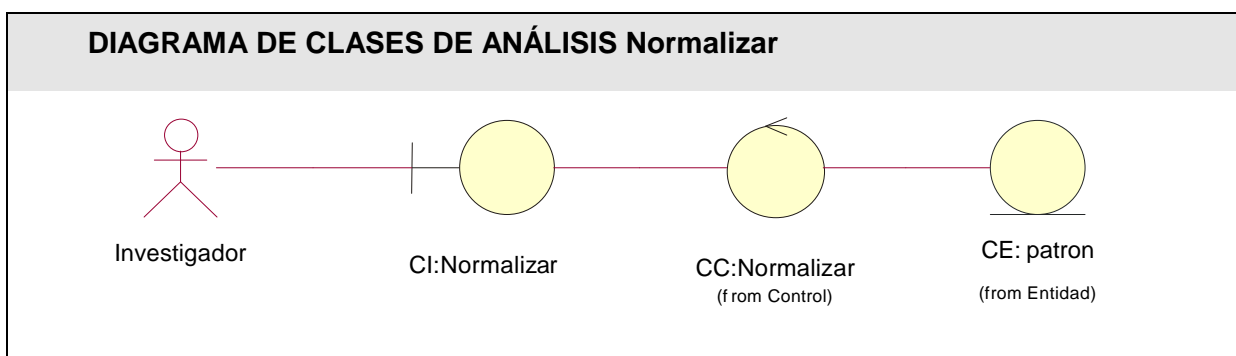


Fig. 5 Diagrama de análisis # 5: Normalizar.

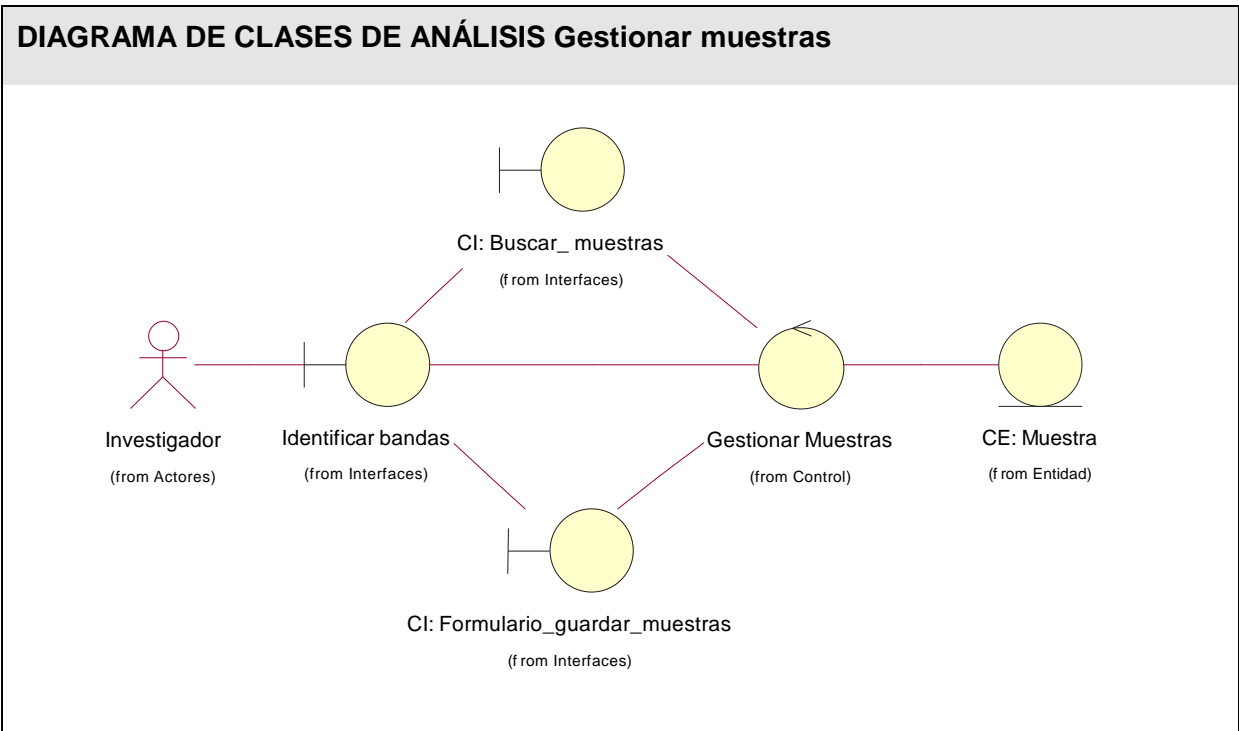


Fig. 6 Diagrama de análisis # 6: Gestionar muestras.

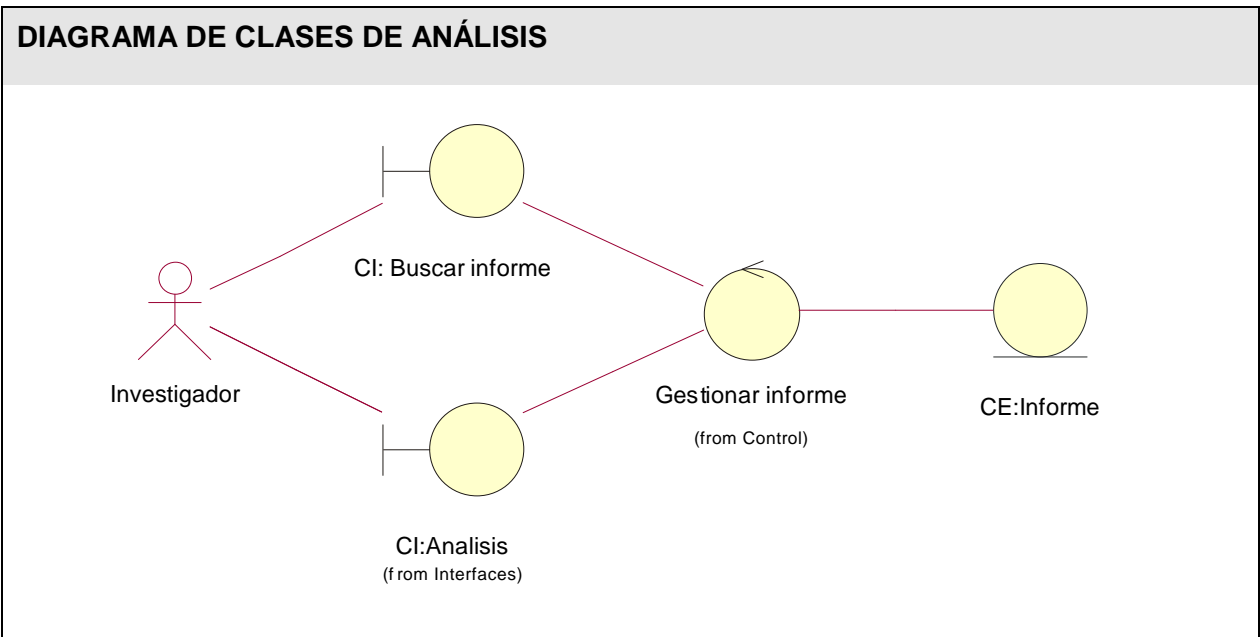


Fig. 7 Diagrama de análisis # 7: Gestionar informe.

## Anexo 3 Diagramas de clases del diseño:

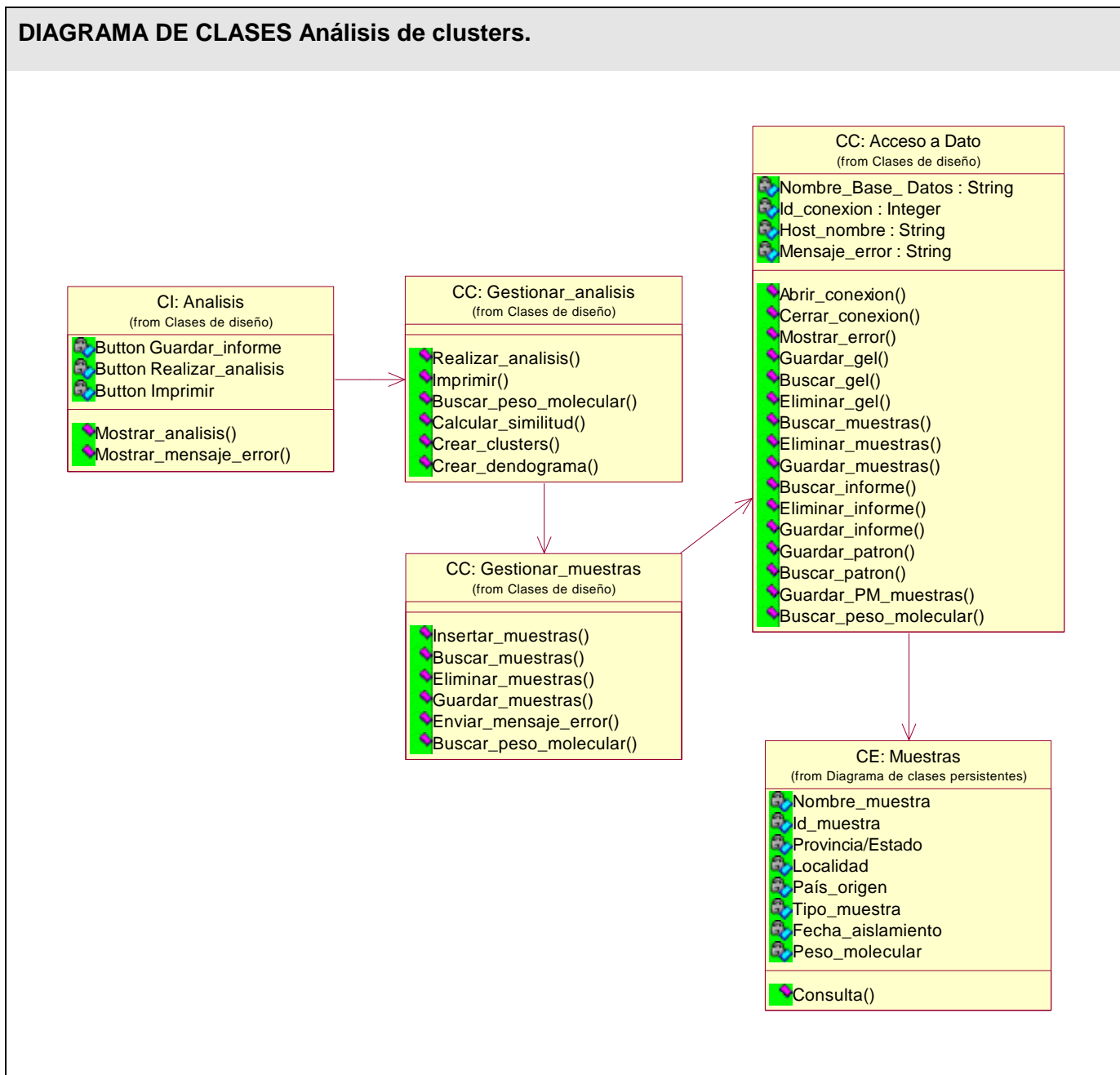


Fig. 8 Diagrama de diseño # 1: Análisis de clusters.



## DIAGRAMA DE CLASES Normalizar gel.

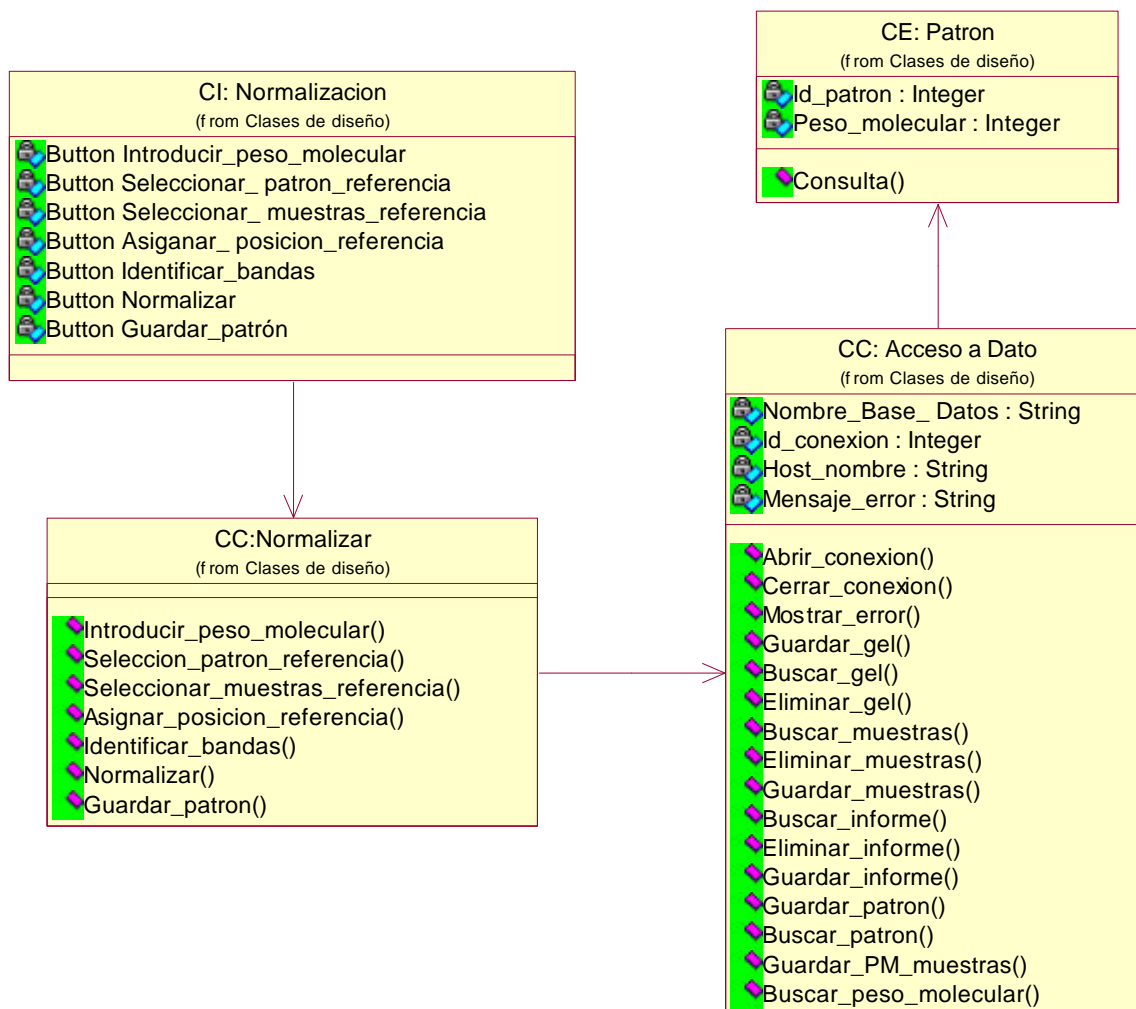
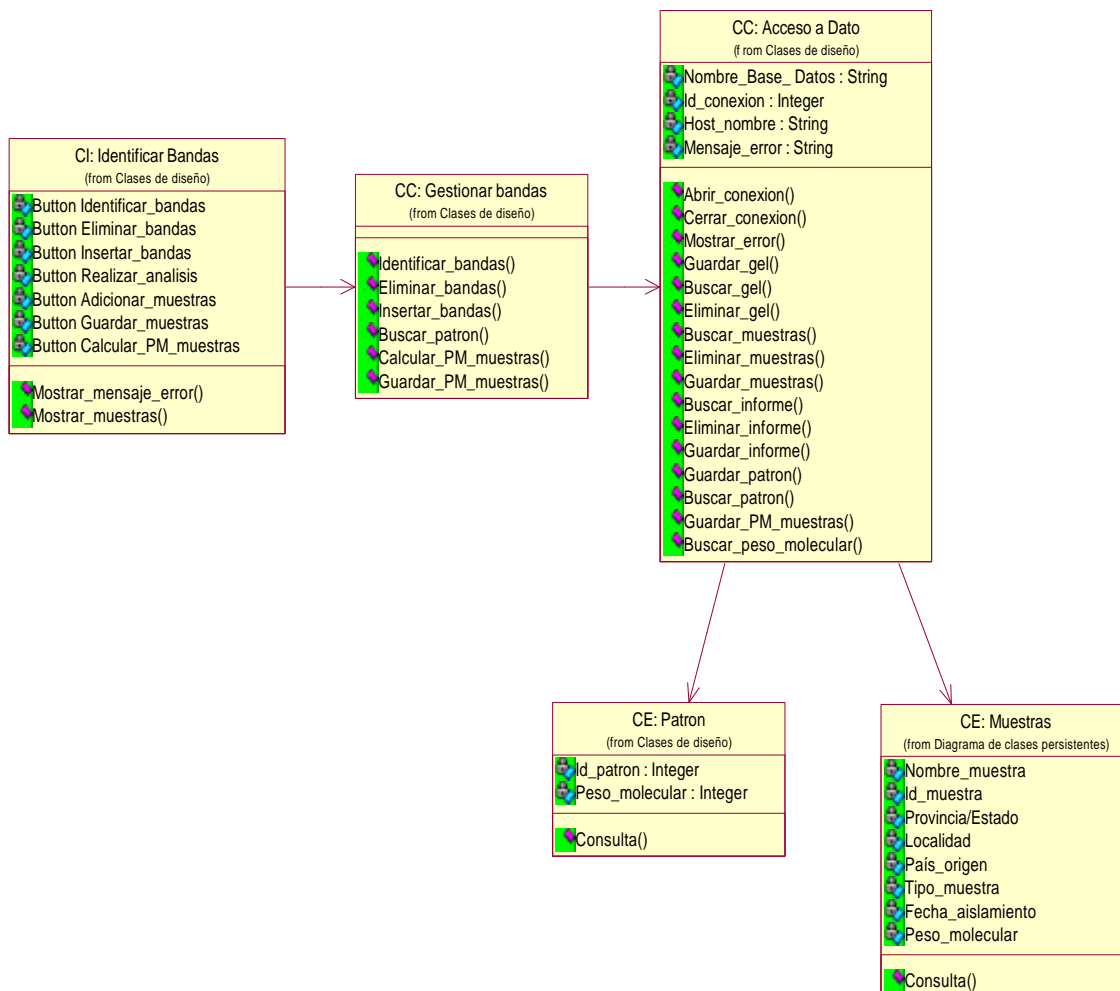


Fig. 9 Diagrama de diseño # 2: Normalizar gel.

**DIAGRAMA DE CLASES Identificar bandas.**



**Fig. 10 Diagrama de diseño # 3: Identificar bandas.**

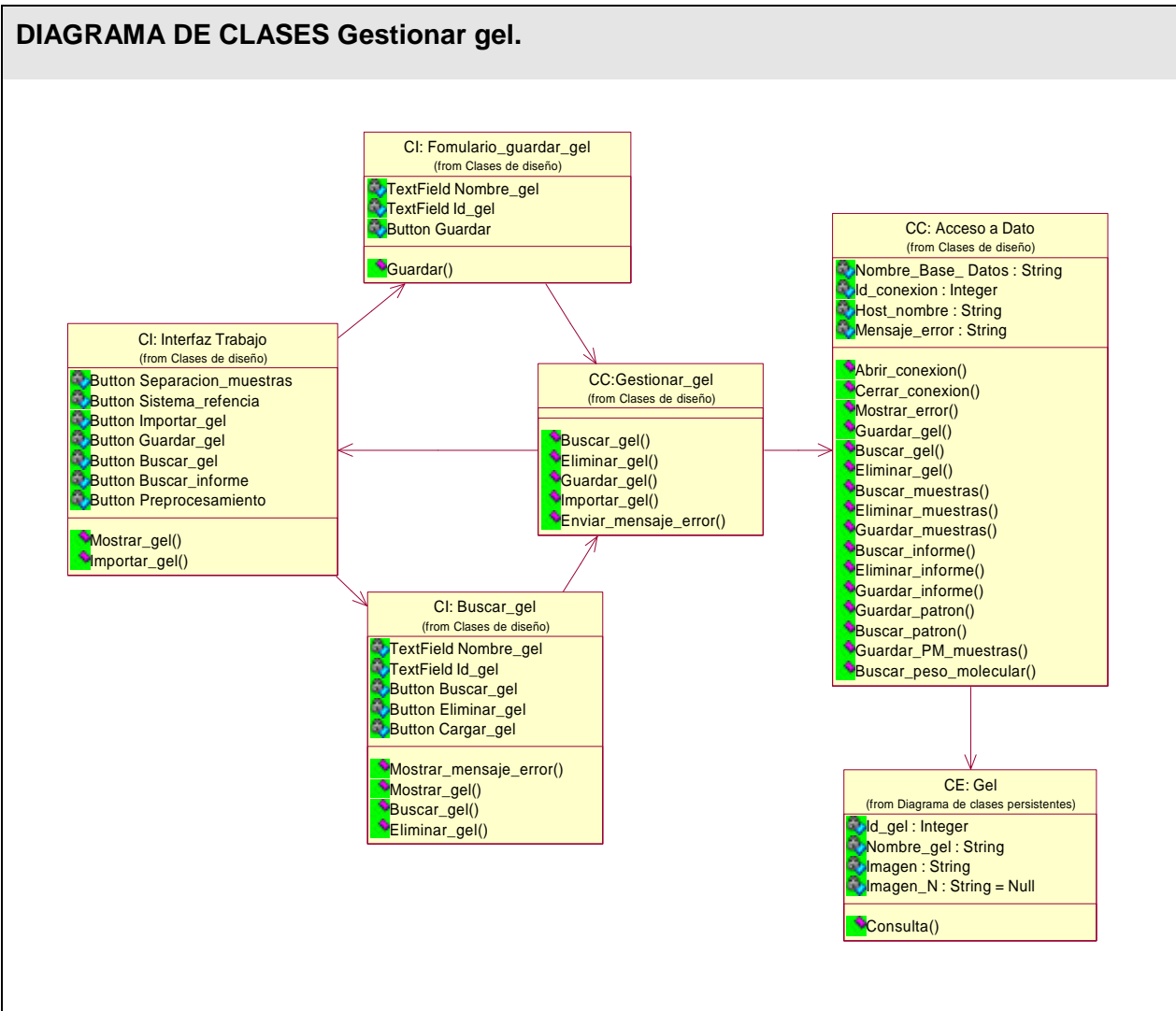
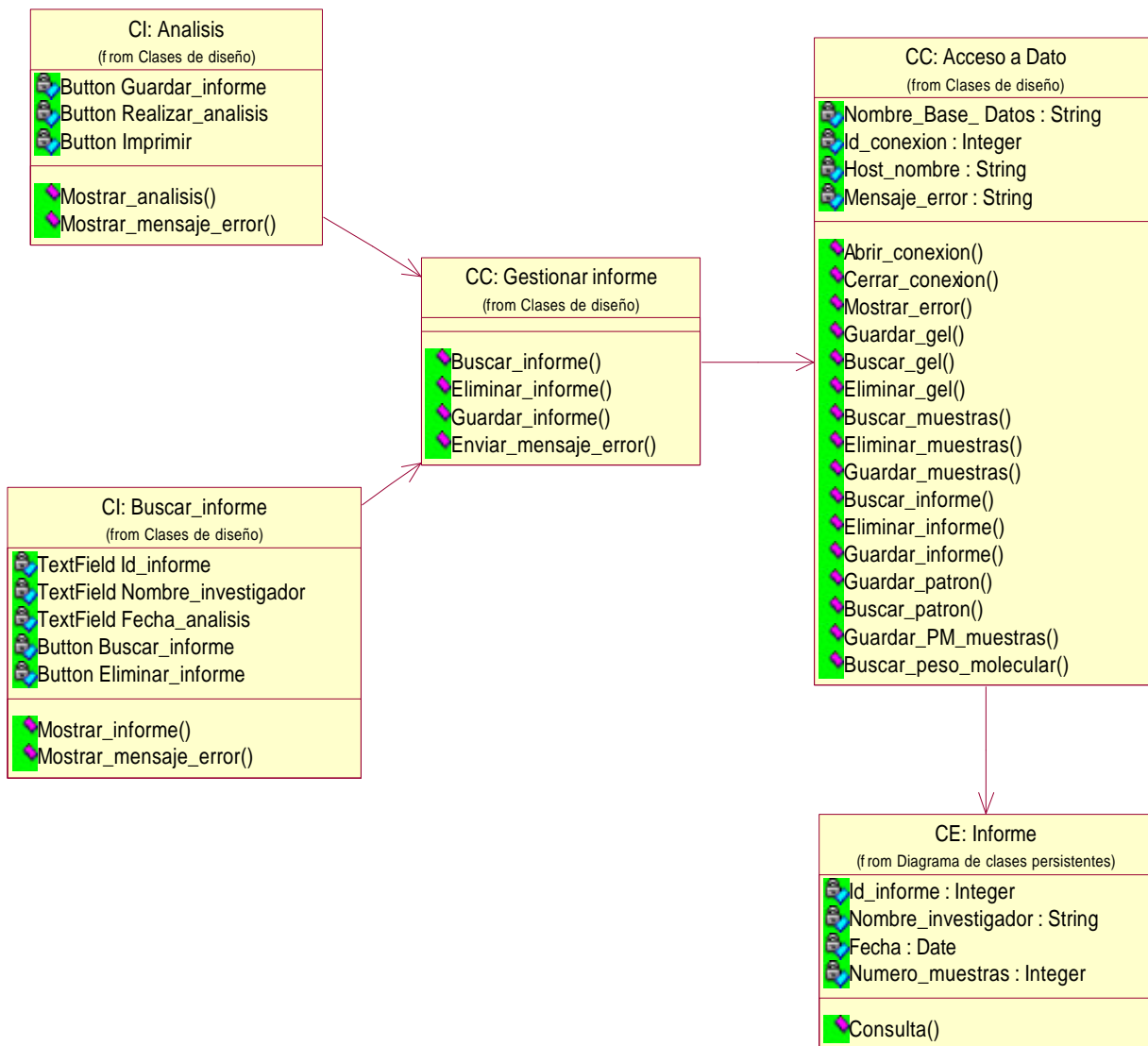


Fig. 11 Diagrama de diseño # 4: Gestionar gel.

**DIAGRAMA DE CLASES Gestionar Informe.**



**Fig. 12 Diagrama de diseño # 5: Gestionar informe.**

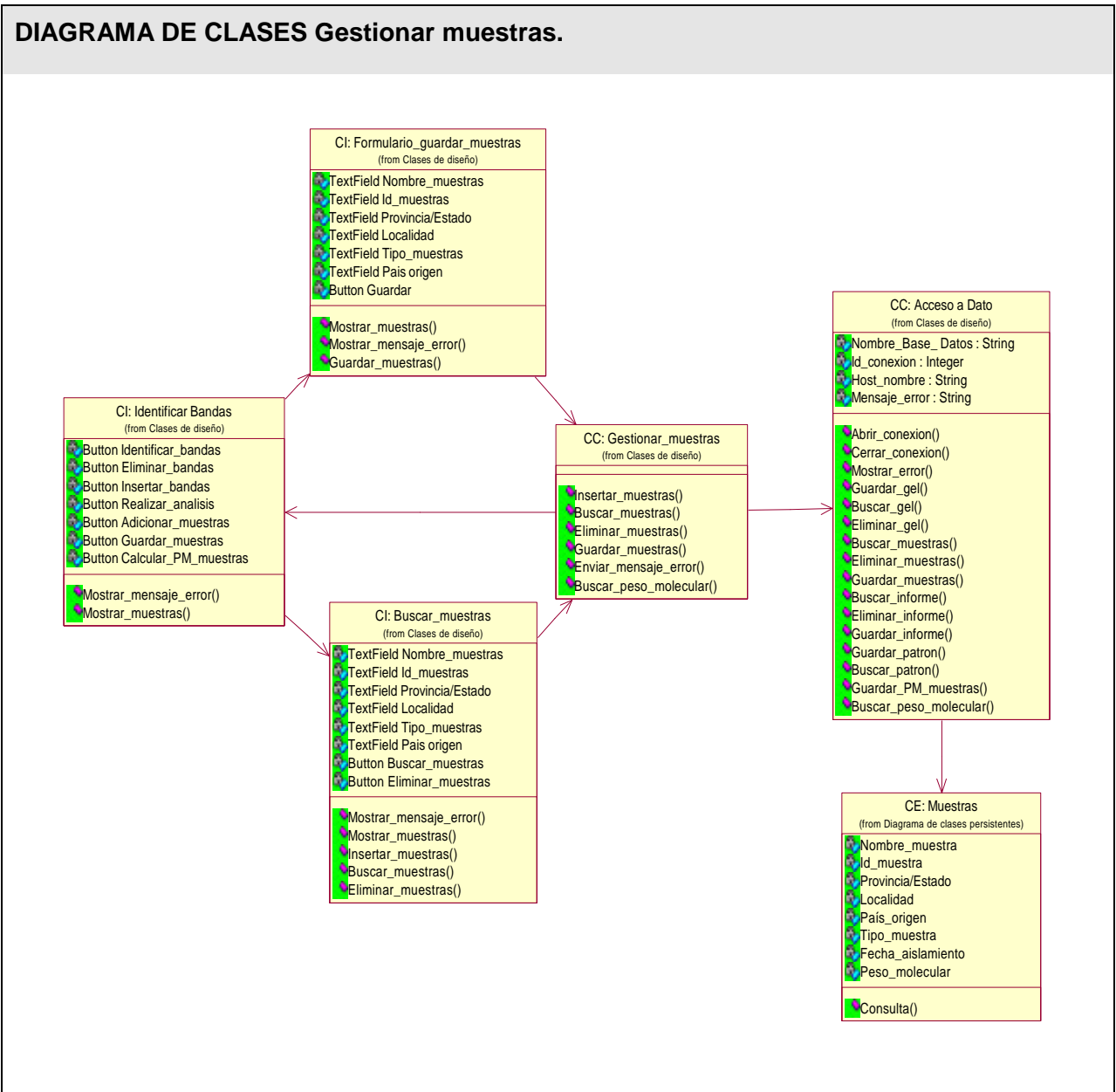
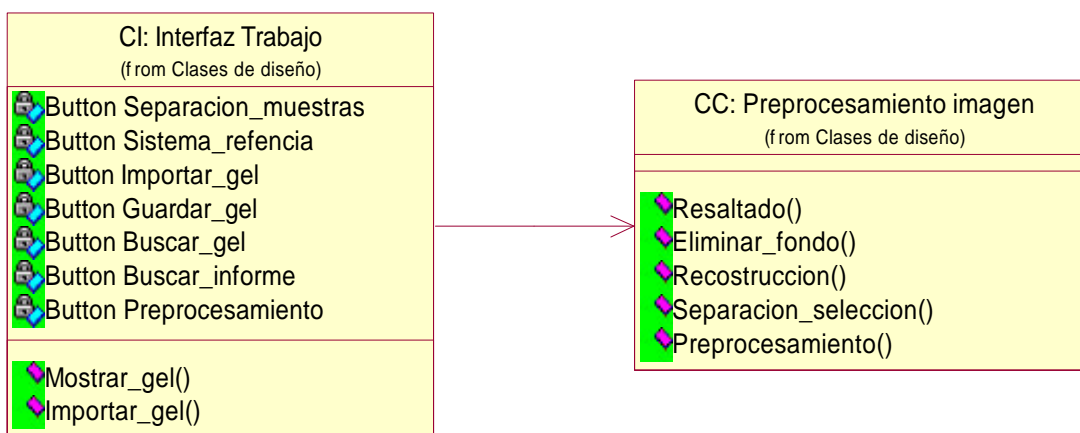


Fig. 13 Diagrama de diseño # 6: Gestionar muestras.

**DIAGRAMA DE CLASES Preprocesamiento.****Fig. 14 Diagrama de diseño # 7: Preprocesamiento.**

## Anexo 4: Diagramas de Interacción.

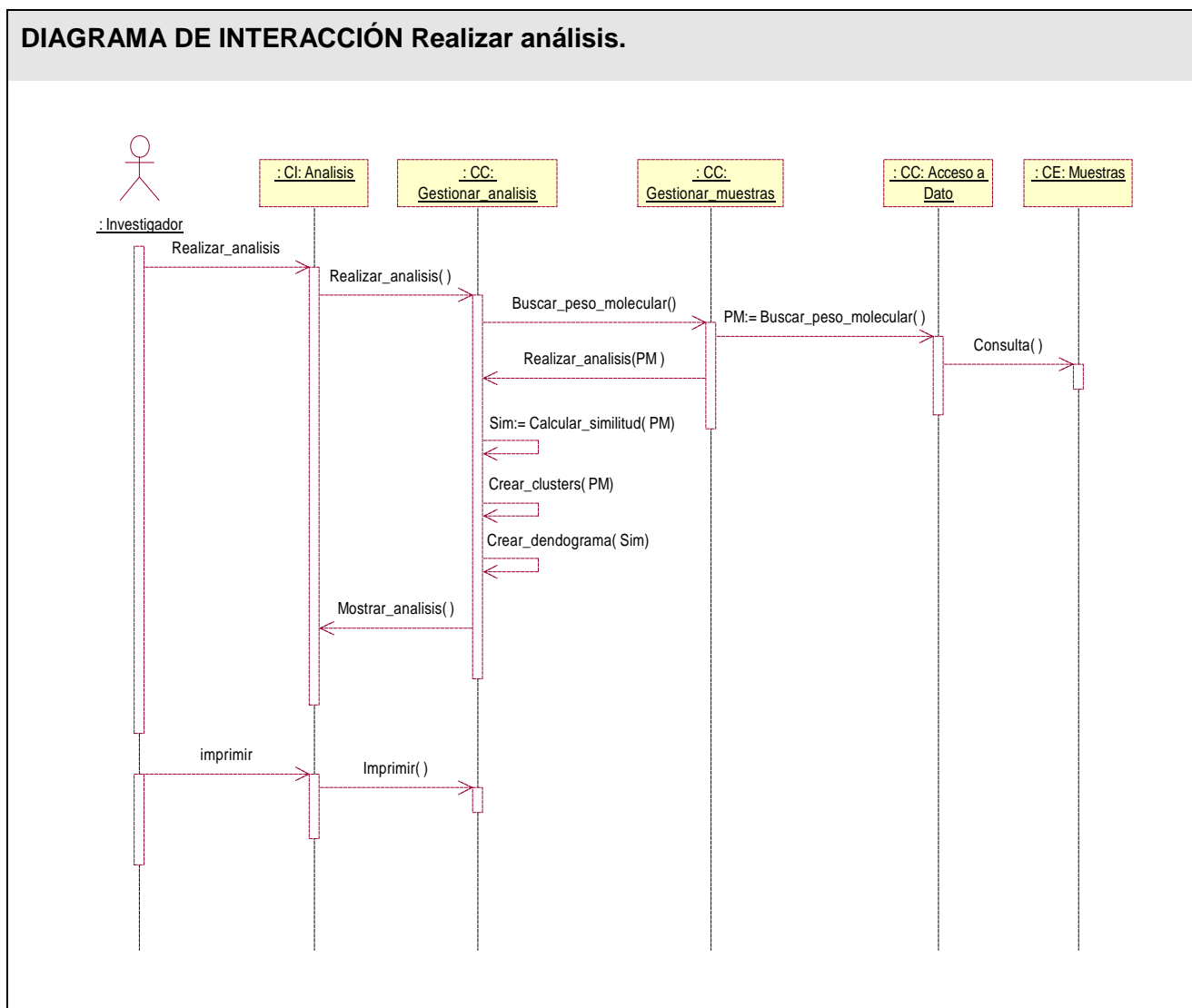


Fig. 15 Diagrama de interacción # 1: Análisis de clusters.

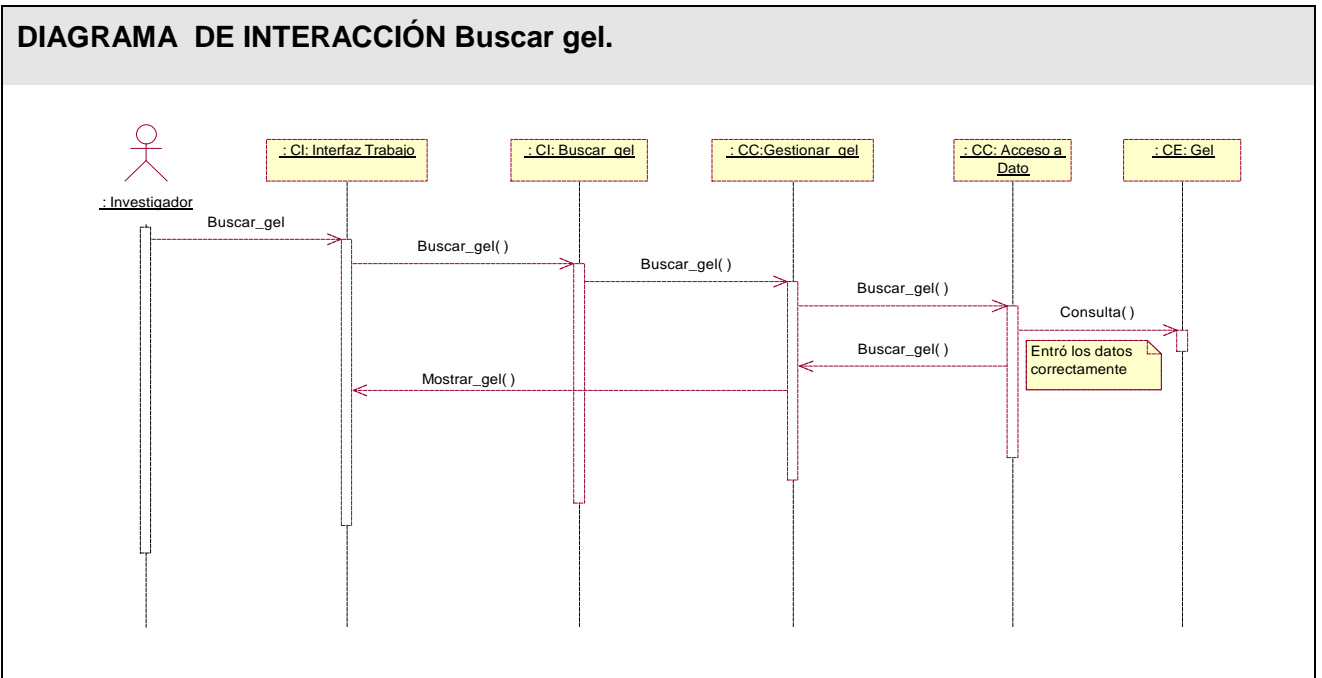


Fig. 16 Diagrama de interacción # 2: Buscar gel.

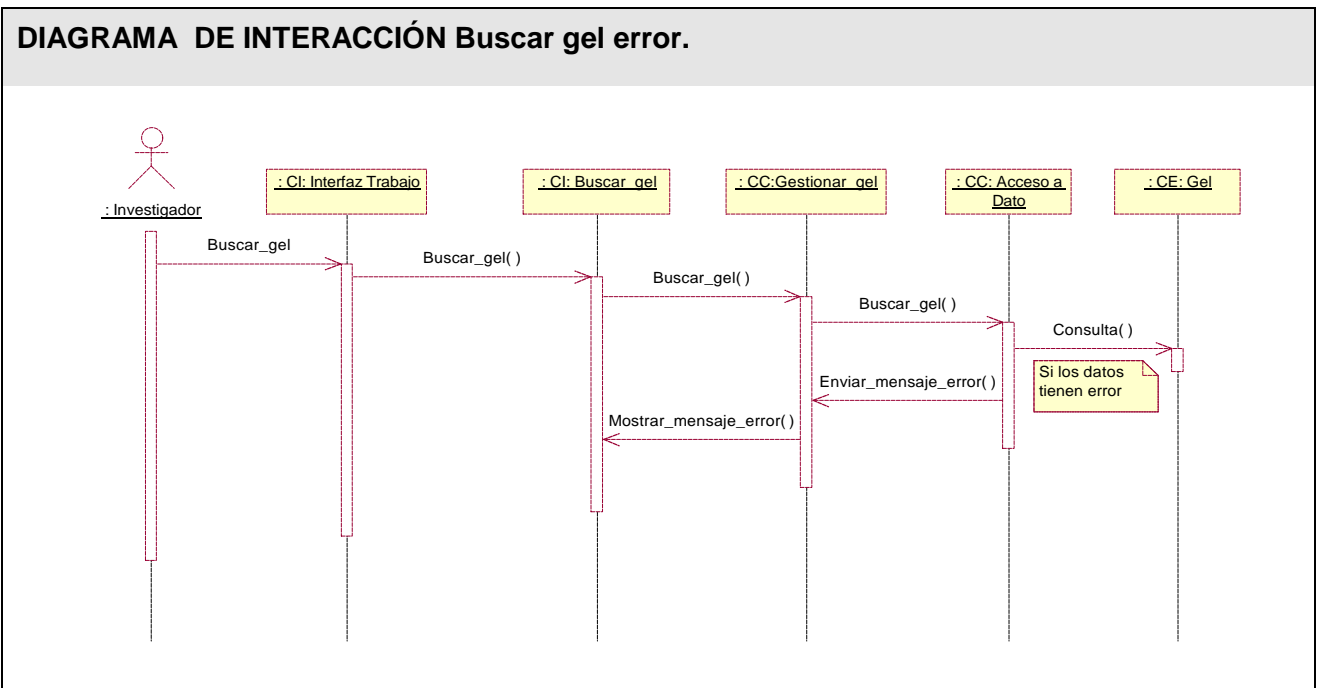


Fig. 17 Diagrama de interacción # 3: Buscar gel error.



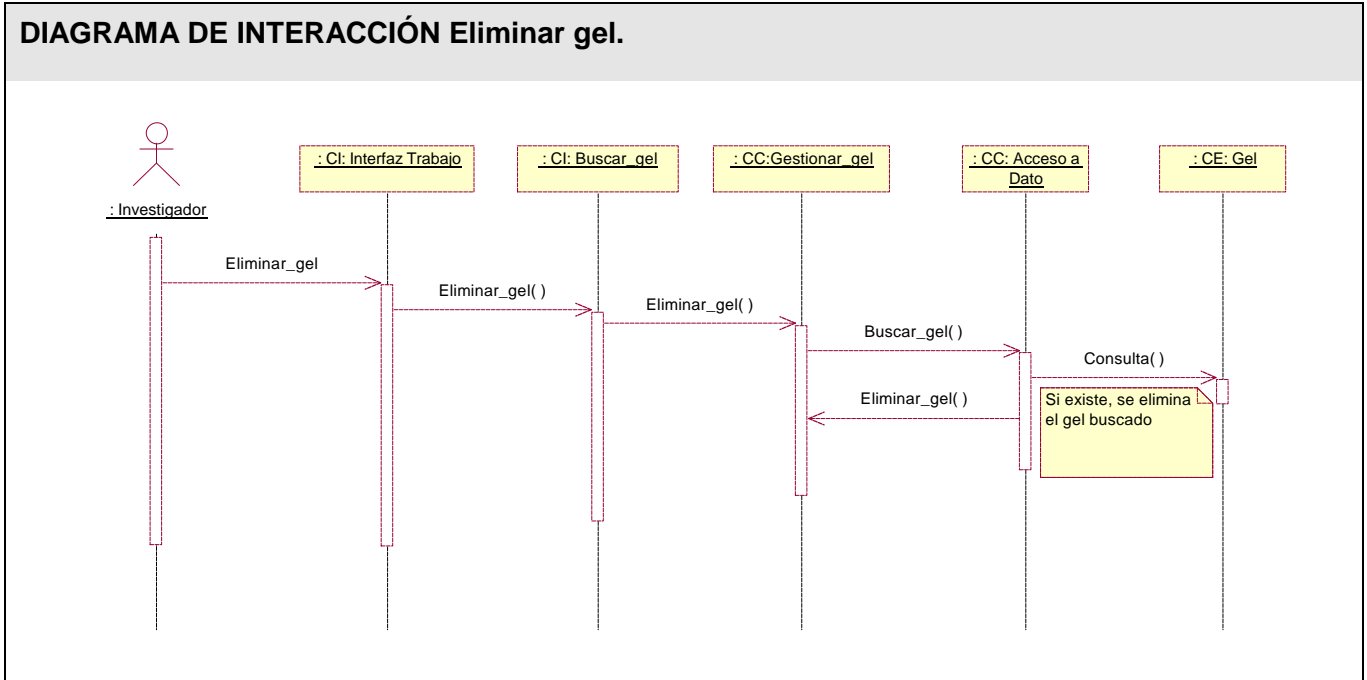


Fig. 18 Diagrama diseño # 4: Eliminar gel.

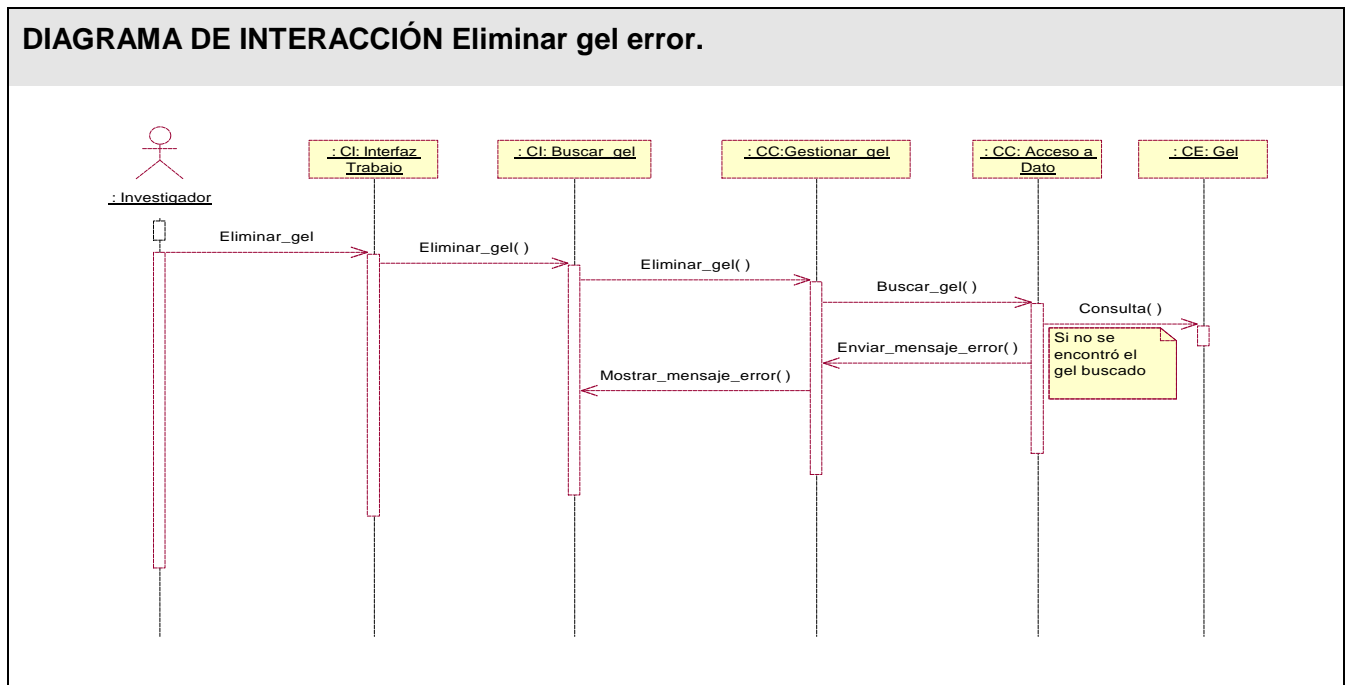


Fig. 19 Diagrama diseño # 5: Eliminar gel error.

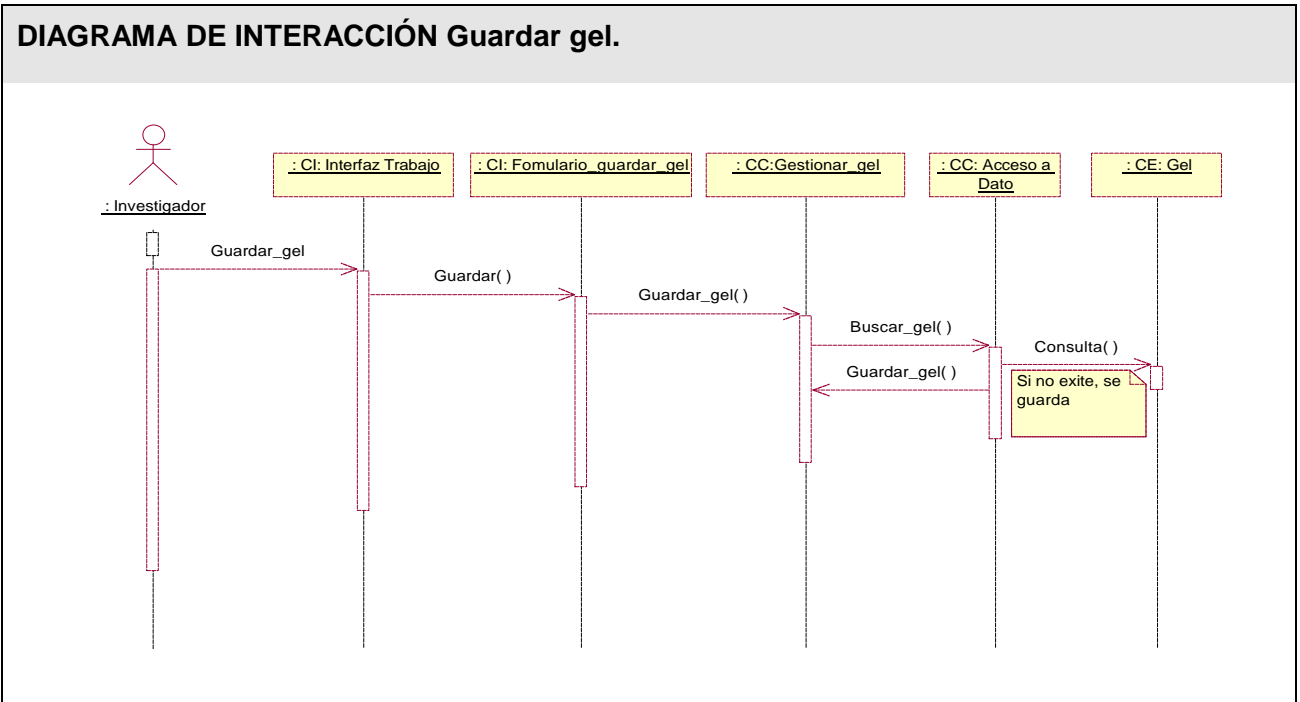


Fig. 20 Diagrama diseño # 6: Guardar gel.

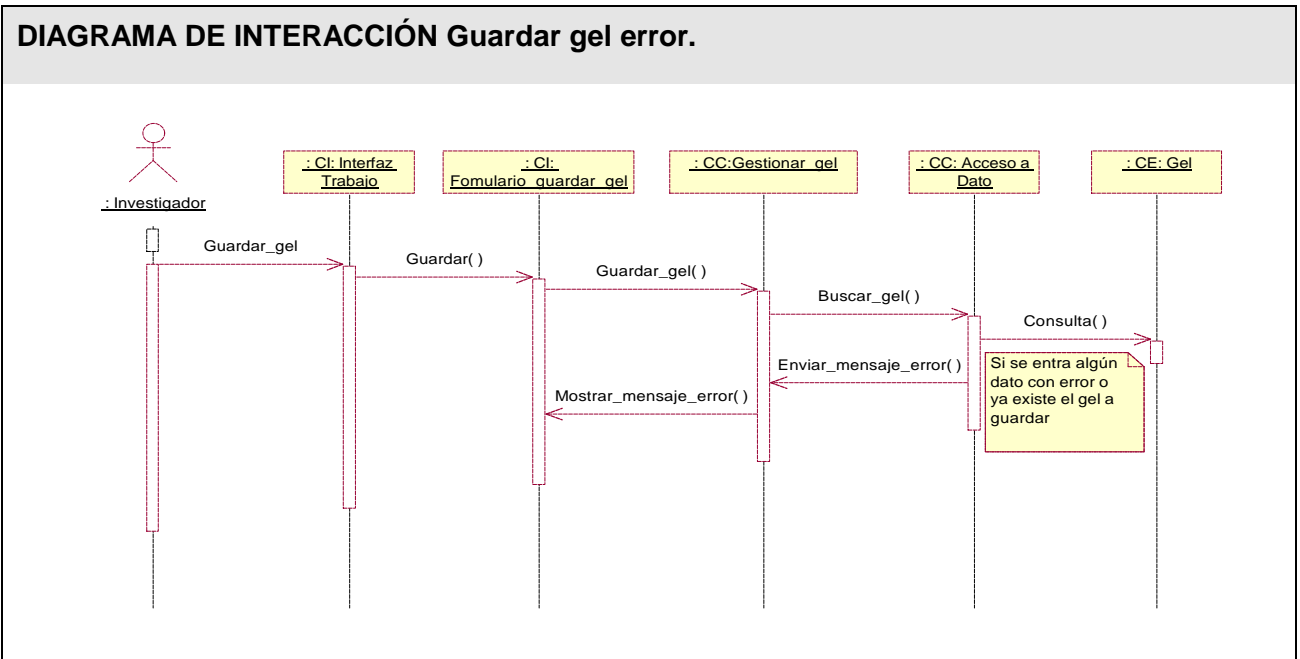


Fig. 21 Diagrama diseño # 7: Guardar gel error.

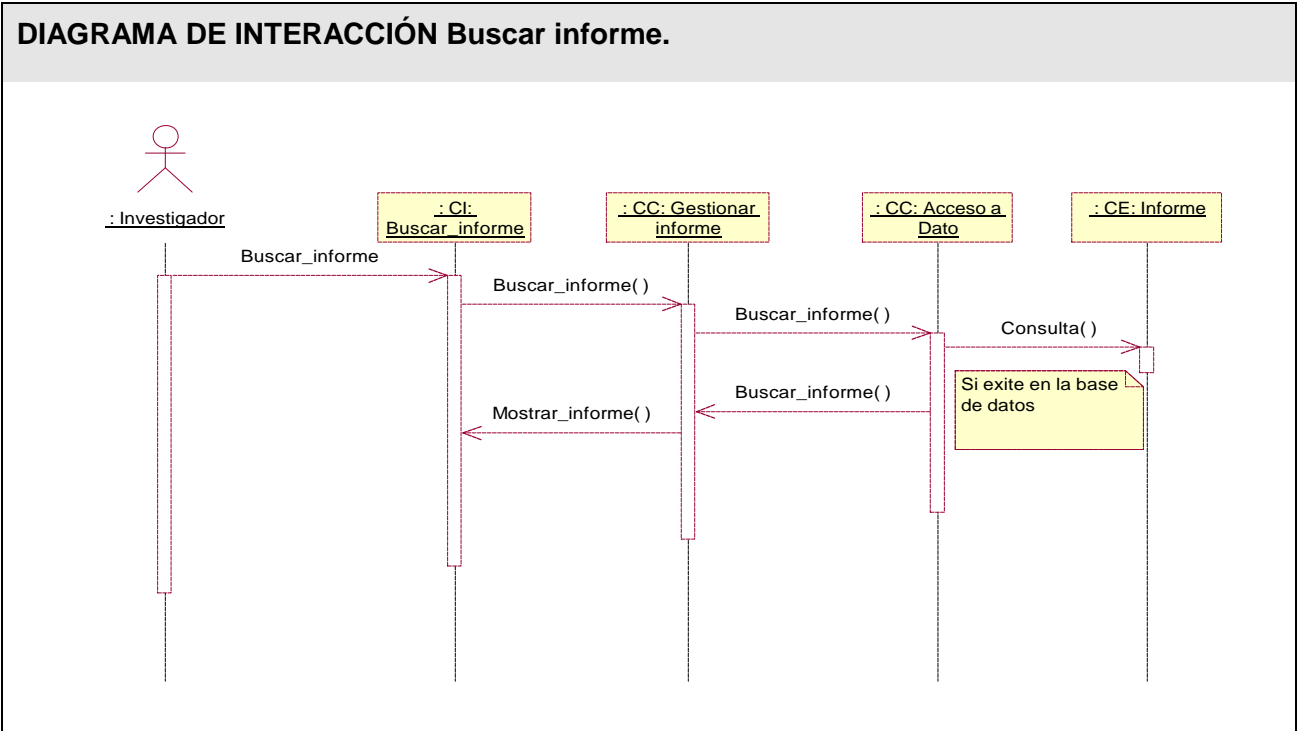


Fig. 22 Diagrama de Interacción # 8: Buscar informe.

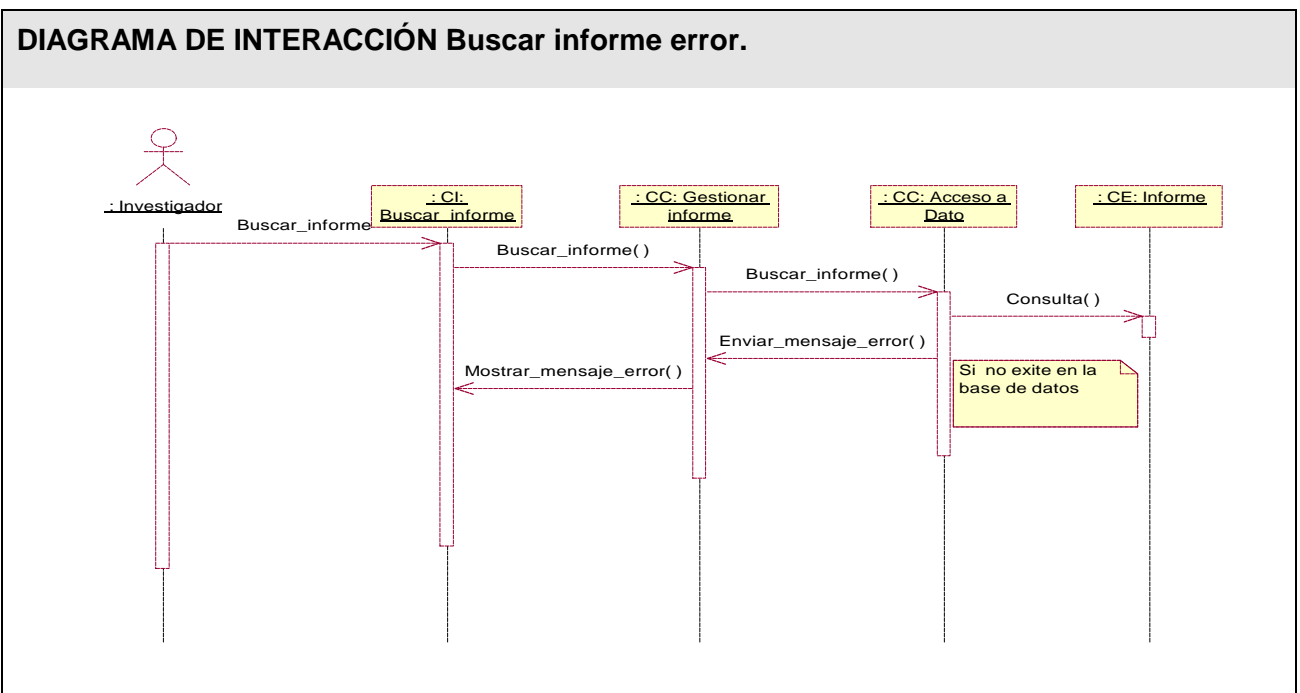


Fig. 23 Diagrama de Interacción # 9: Buscar informe error.

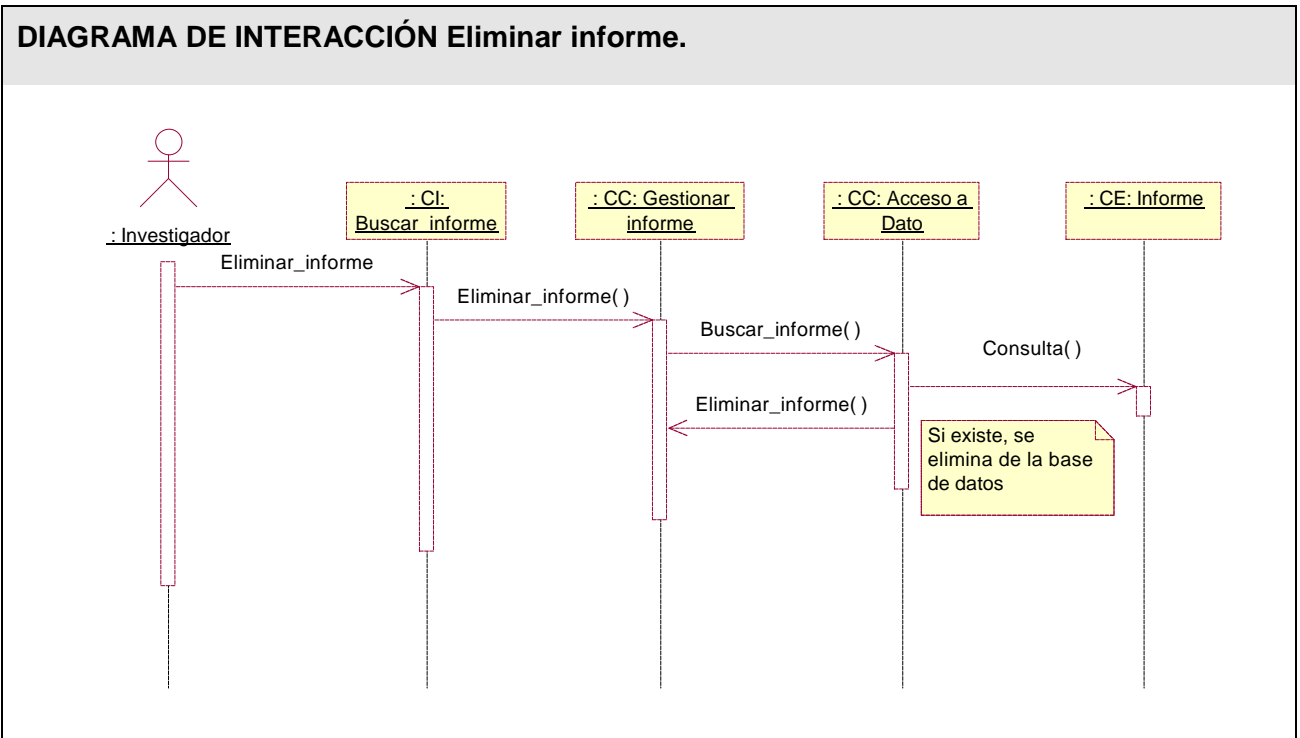


Fig. 24 Diagrama de Interacción # 10: Eliminar informe.

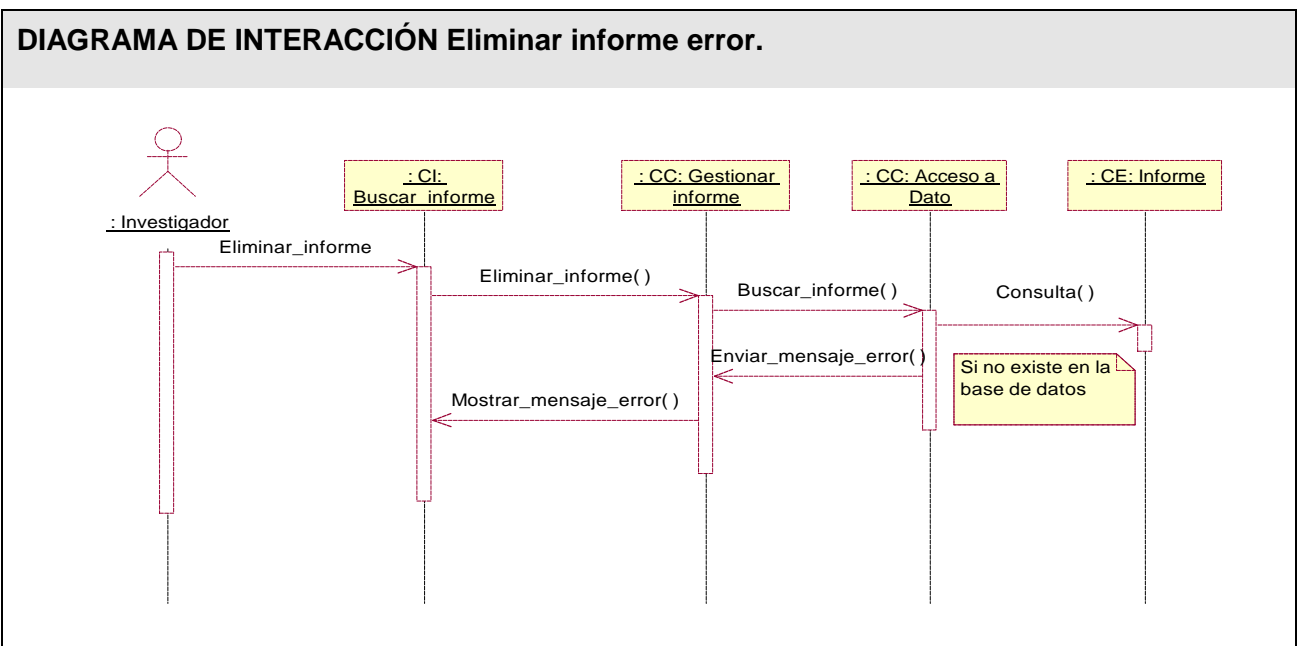


Fig. 25 Diagrama de Interacción # 11: Eliminar informe error.



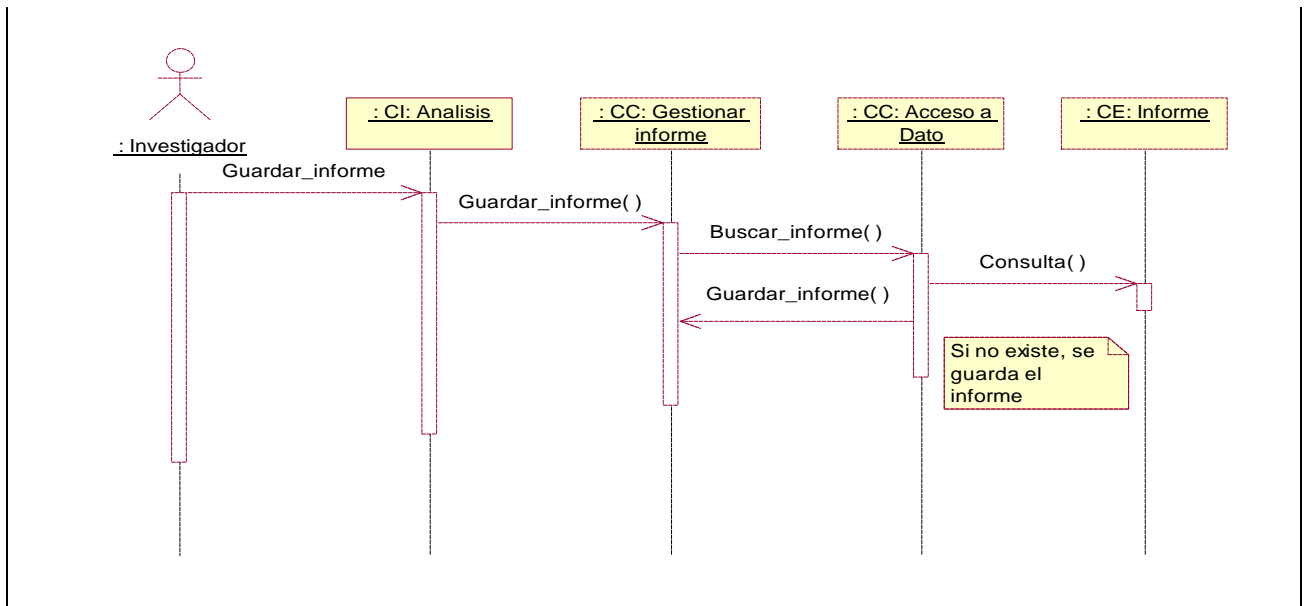


Fig. 26 Diagrama de Interacción # 12: Guardar informe.

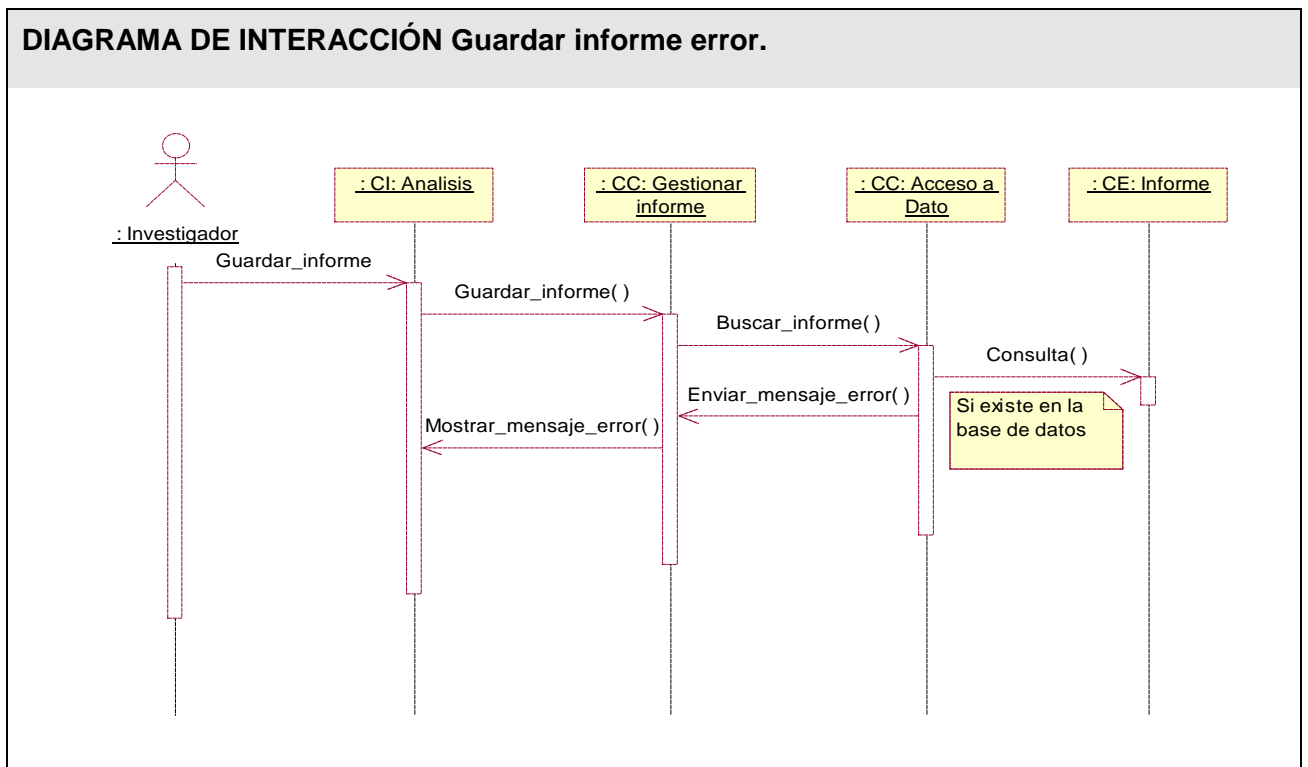


Fig. 27 Diagrama de Interacción # 13: Guardar informe error.

**DIAGRAMA DE INTERACCIÓN Buscar muestras.**

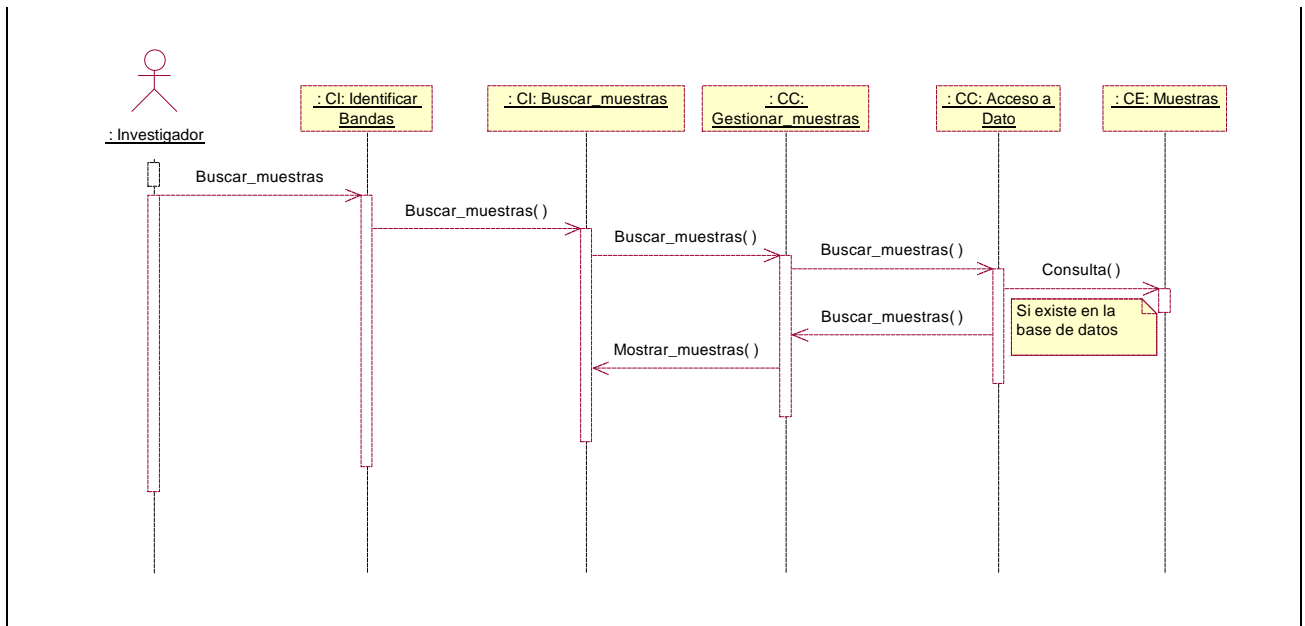


Fig. 28 Diagrama de Interacción # 14: Buscar muestras.

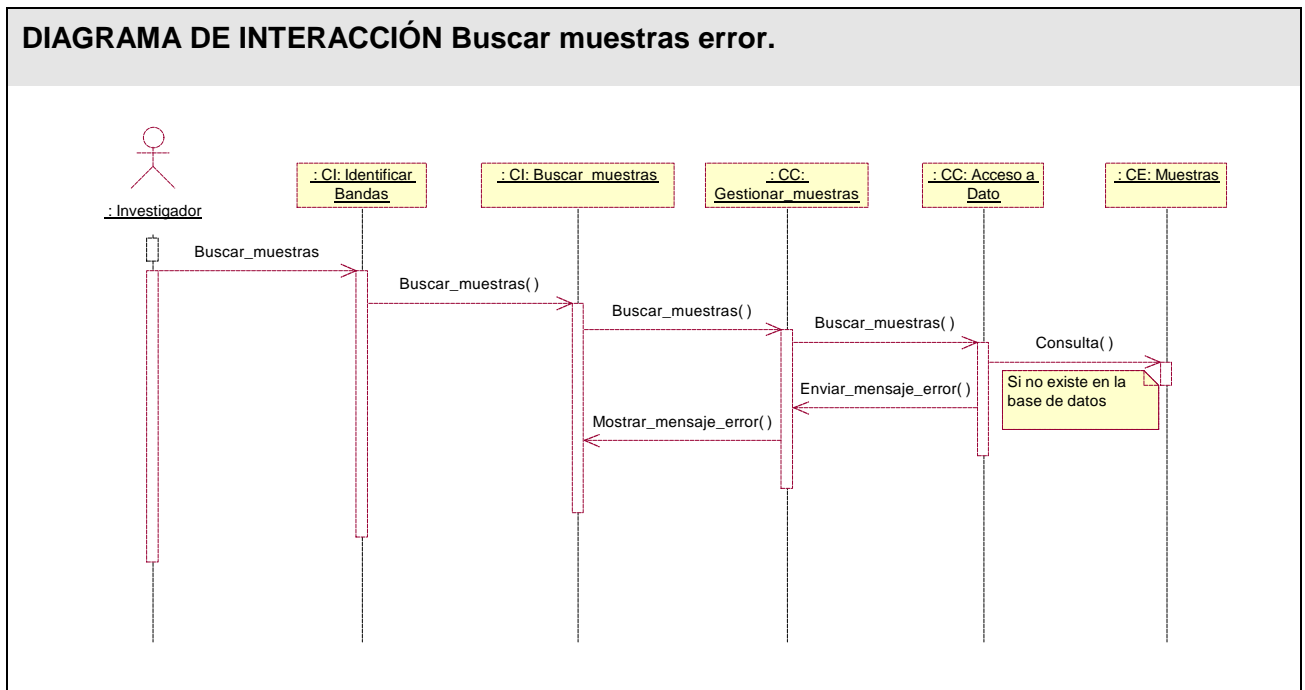


Fig. 29 Diagrama de Interacción # 15: Buscar muestras error.

**DIAGRAMA DE INTERACCIÓN Eliminar muestras**

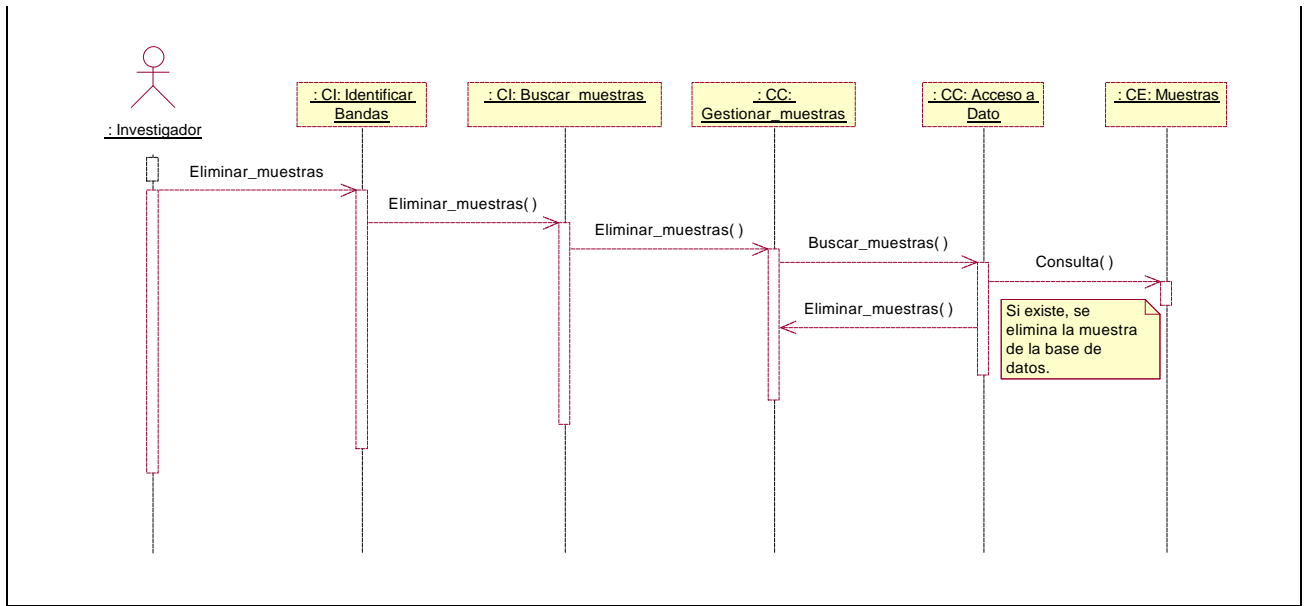


Fig. 30 Diagrama de Interacción # 16: Eliminar muestras.

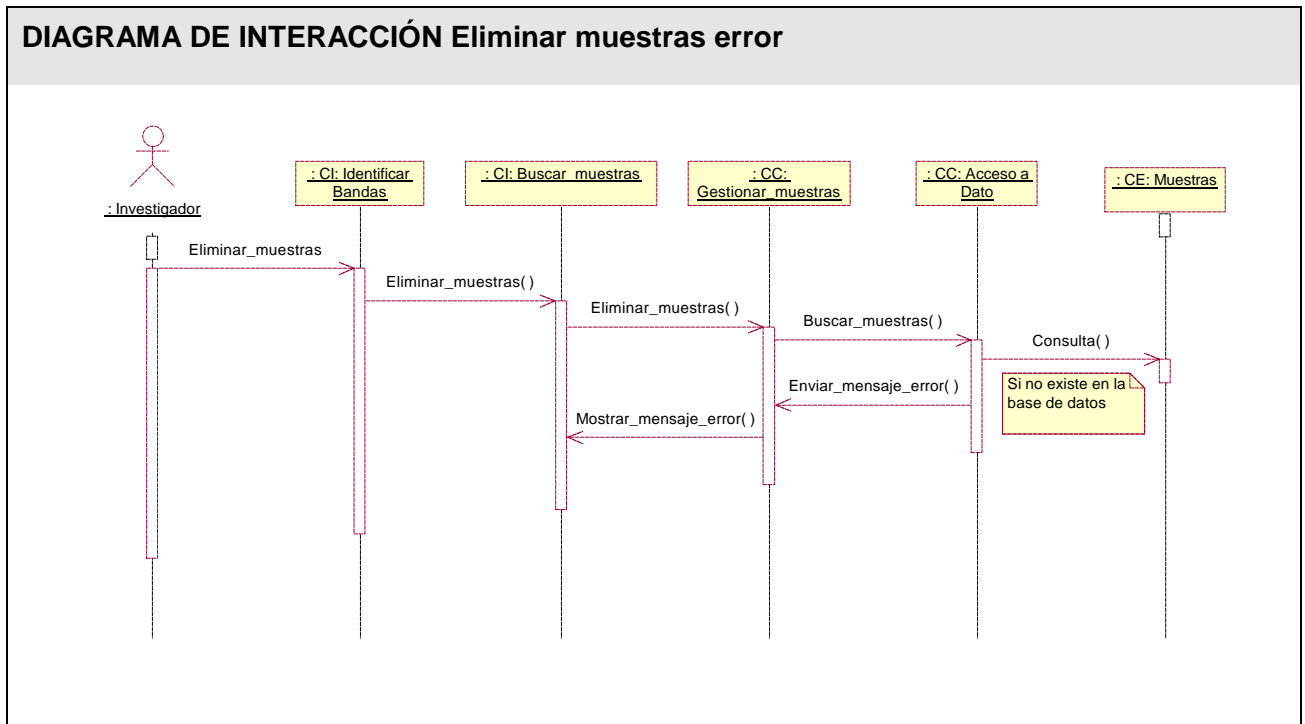


Fig. 31 Diagrama de Interacción # 17: Eliminar muestras error.

**DIAGRAMA DE INTERACCIÓN Guardar muestras**

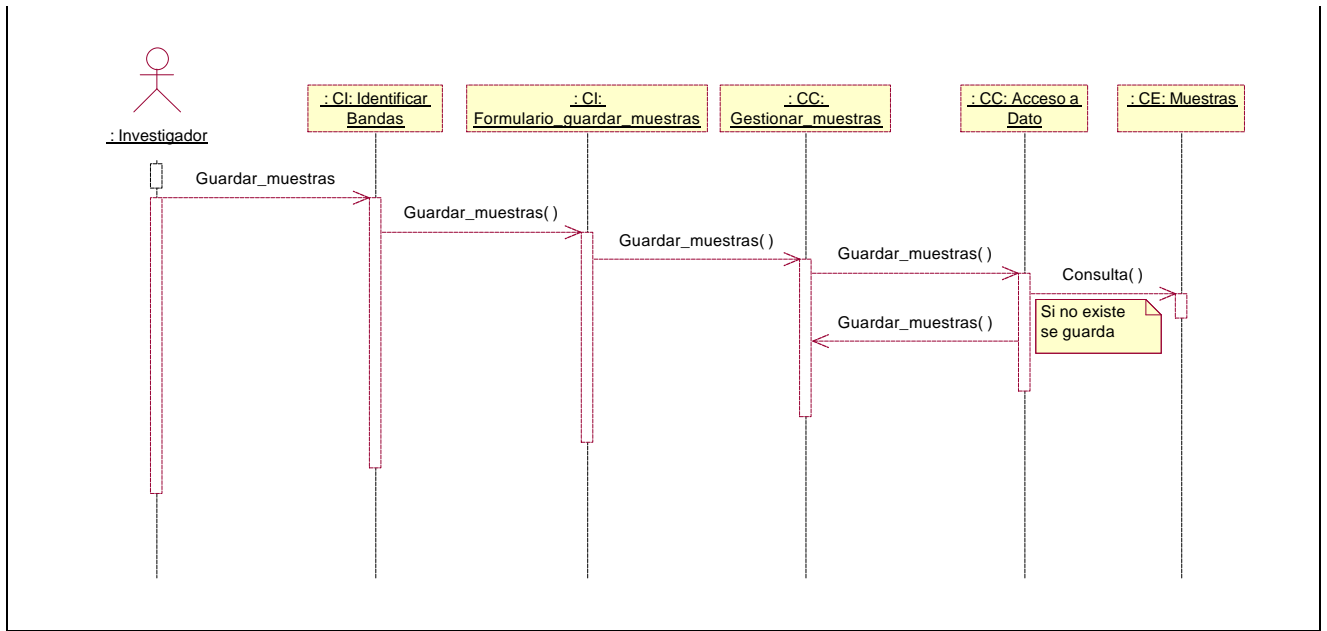


Fig. 32 Diagrama de Interacción # 18: Guardar muestras.

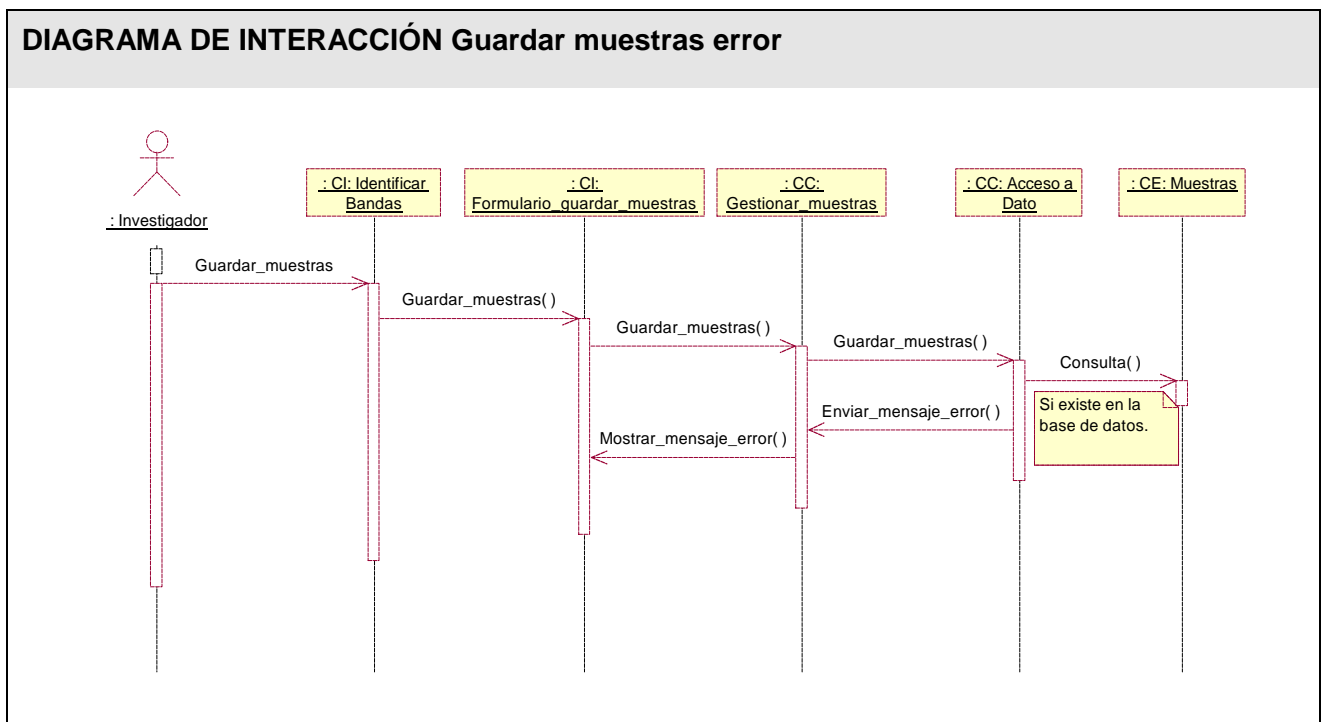


Fig. 33 Diagrama de Interacción # 19: Guardar muestras error.

**DIAGRAMA DE INTERACCIÓN Insertar muestras**



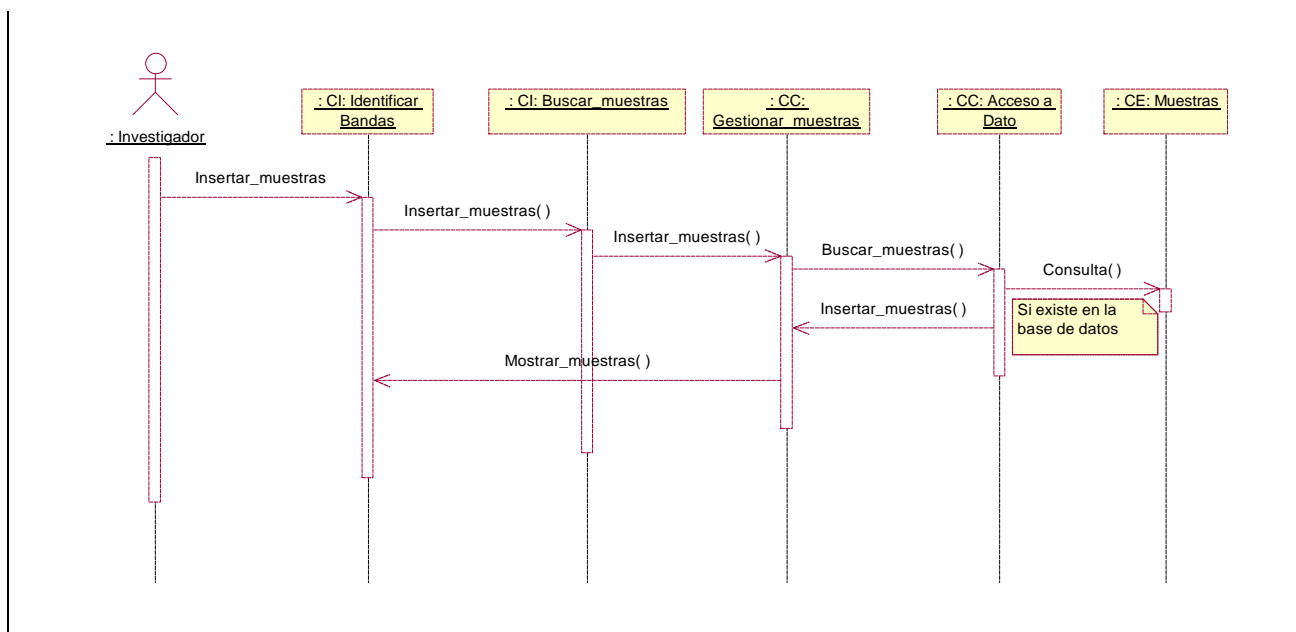


Fig. 34 Diagrama de Interacción # 20: Insertar muestras.

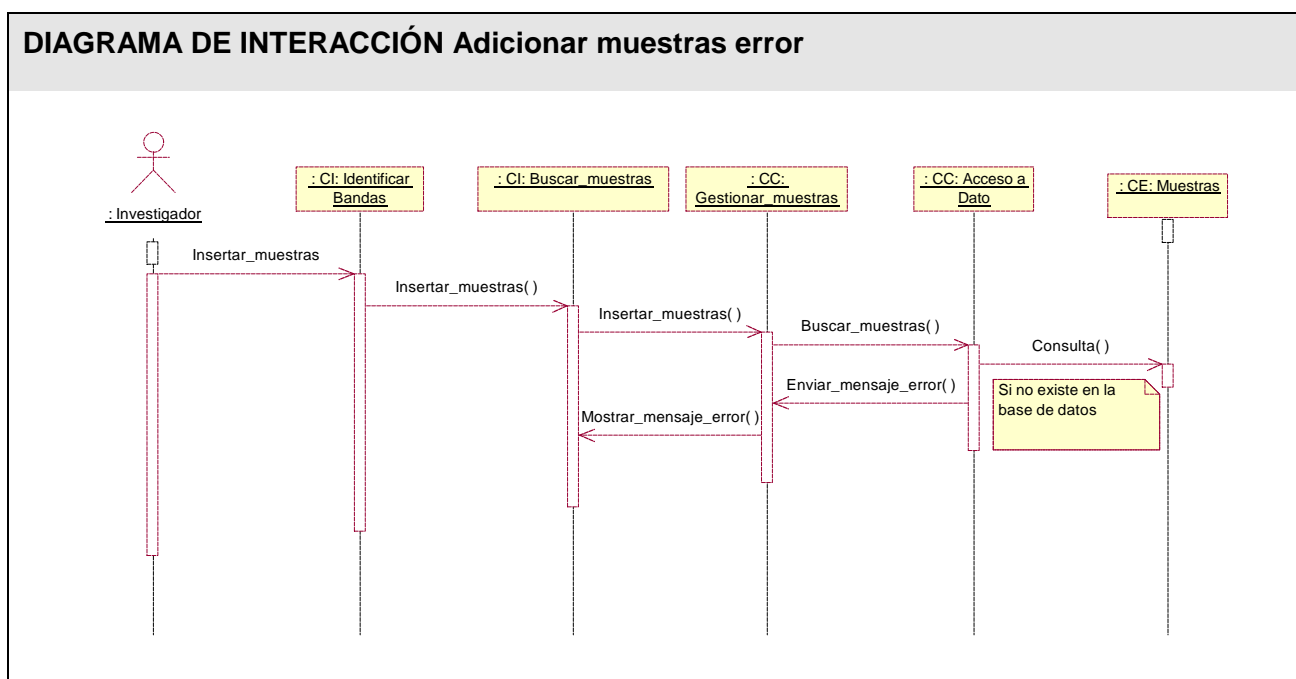


Fig. 35 Diagrama de Interacción # 21: Adicionar muestras error.

**DIAGRAMA DE INTERACCIÓN Calcular peso molecular.**

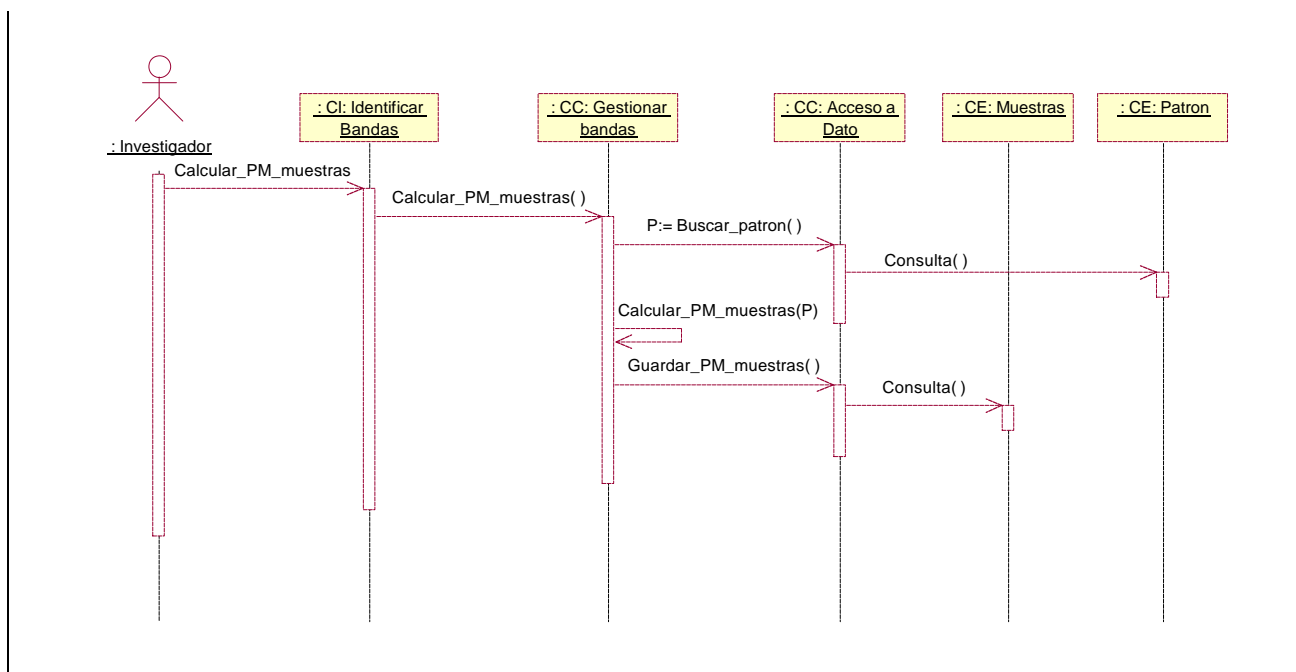


Fig. 36 Diagrama de Interacción # 22: Calcular peso molecular.

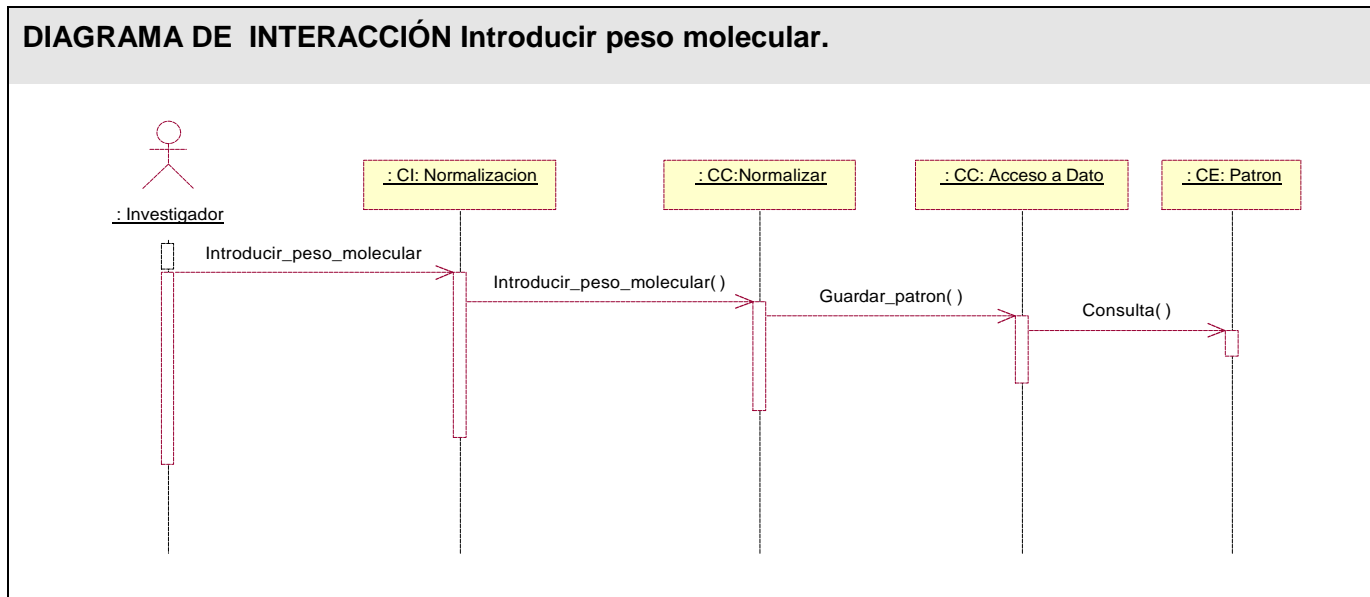


Fig. 37 Diagrama de Interacción # 23: Introducir peso molecular.

## Anexo 4 Diagrama de clases persistentes:

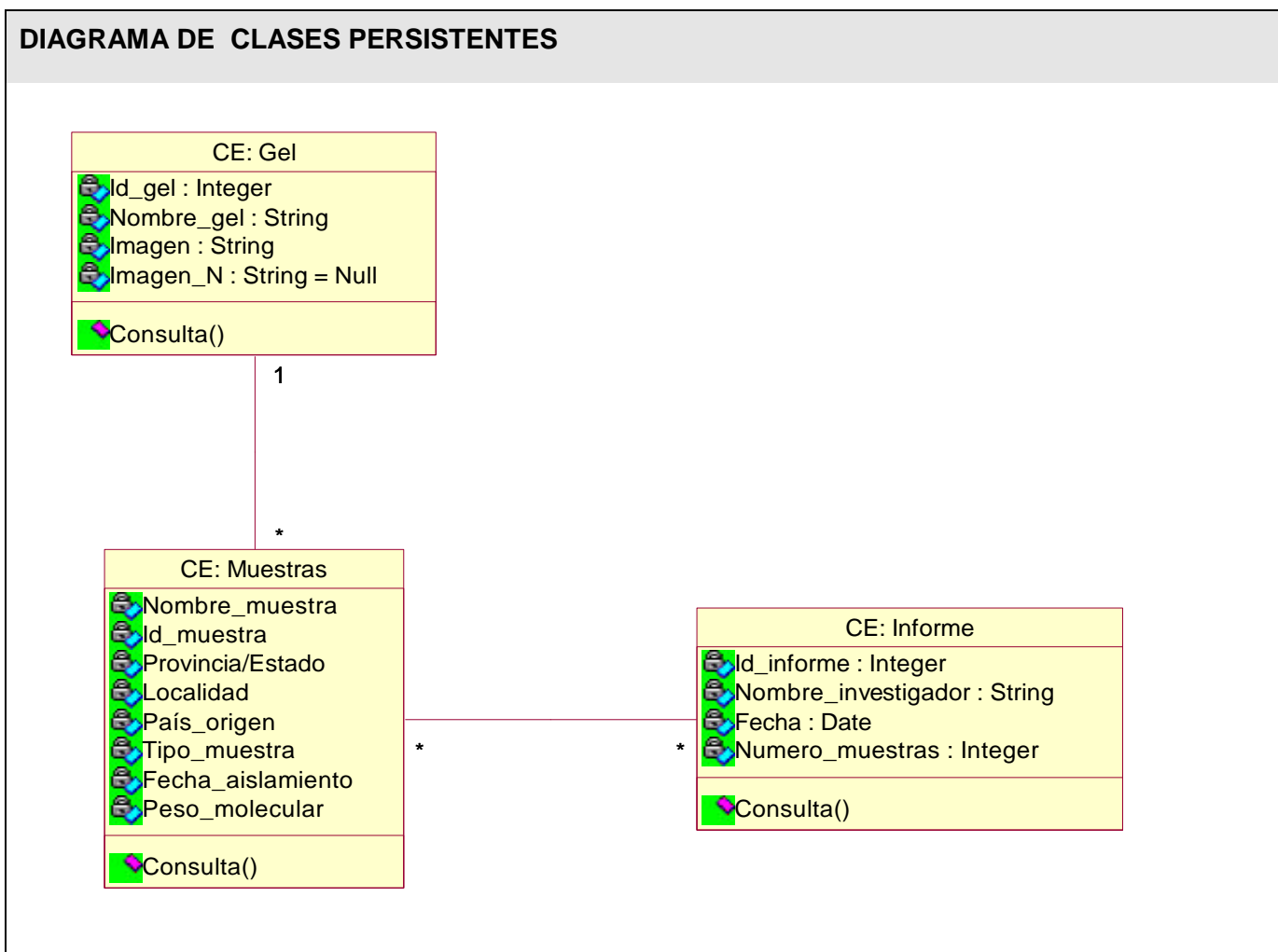


Fig. 38 Diagrama de clases persistentes.

Anexo 5 Diagrama del Modelo de Datos:

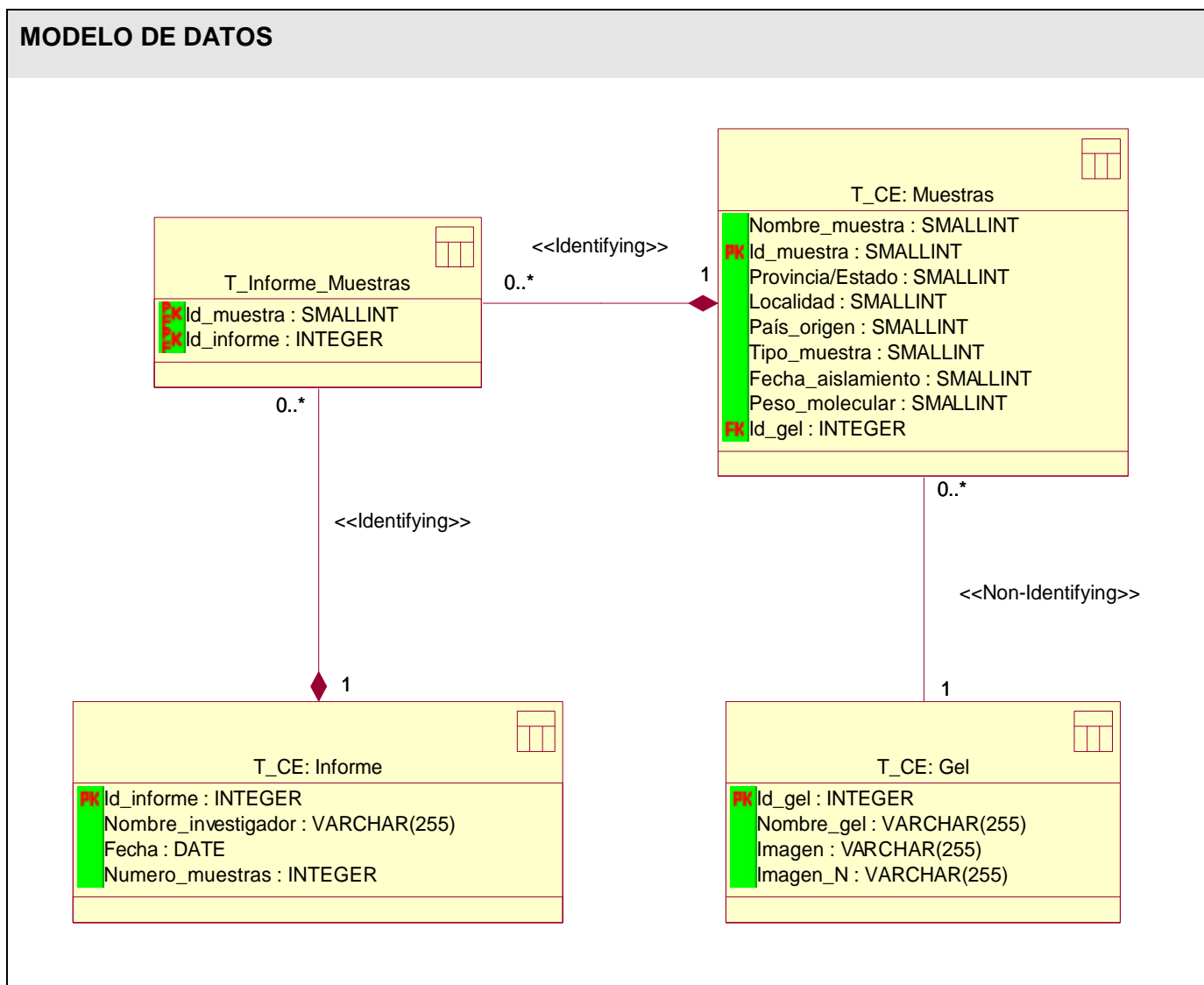


Fig. 39 Diagrama de modelo de datos.

Anexo 6 Diagrama de despliegue:

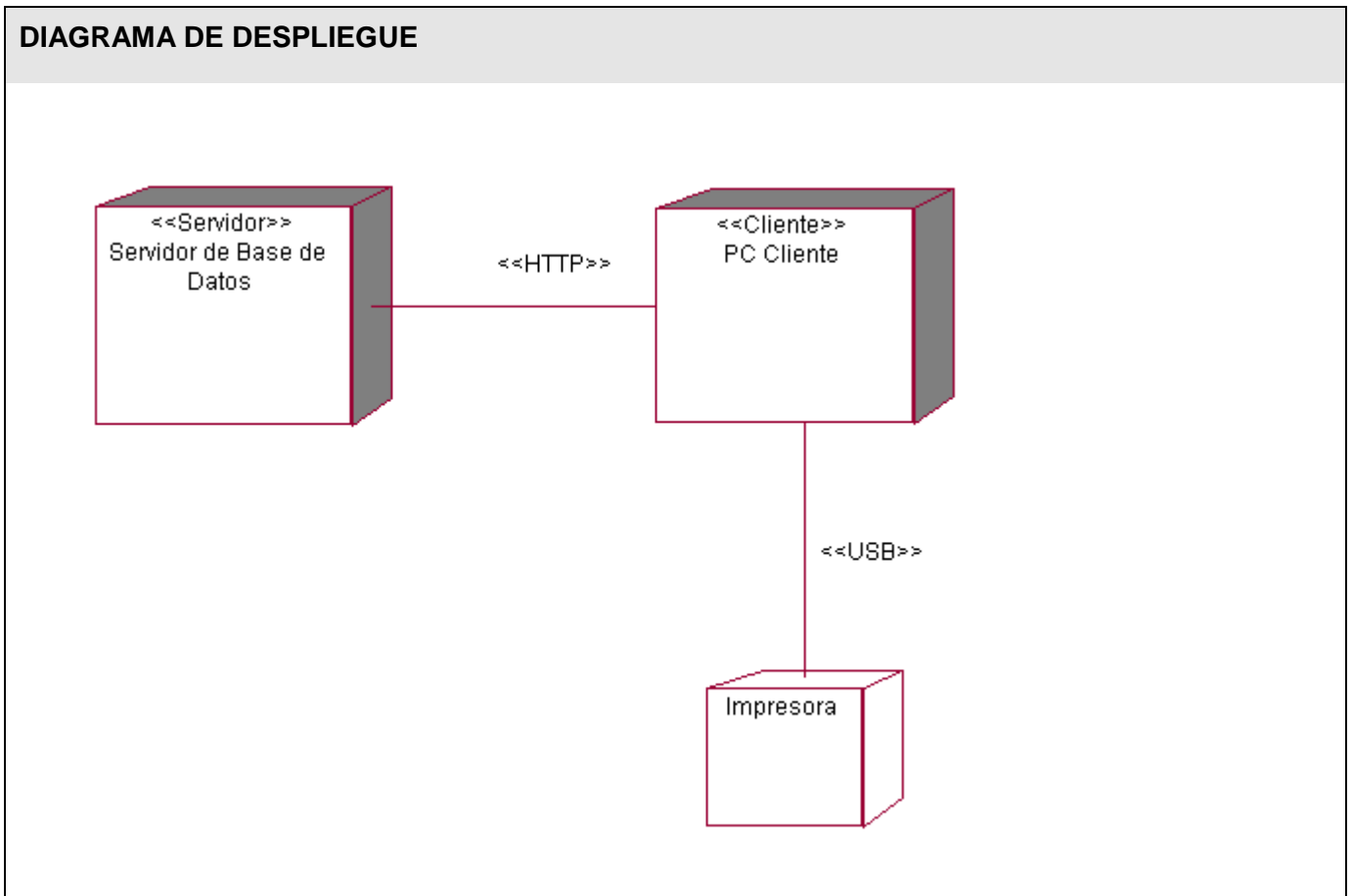


Fig. 40 Diagrama de despliegue.

## Glosario de Términos

### A

**ADN: (Acido desoxirribonucleico)** Constituye el principal componente del material genético de la inmensa mayoría de los organismos, junto con el ARN. Es el componente químico primario de los cromosomas y el material en el que los genes están codificados.

### B

**Base de Datos:** Conjunto de datos almacenados en un soporte informático no volátil. En ella los datos se encuentran interrelacionados y estructurados de acuerdo con su modelo capaz de recoger el máximo de contenido. Se organizan en tablas.

### C

**Cluster:** Es que un conjunto de técnicas que se utilizan para clasificar los objetos o casos en grupos relativamente homogéneos llamados conglomerados.

**Clase de Interfaz:** Modelan la interacción entre el sistema y sus actores (usuarios y sistemas externos), reciben y presentan información y peticiones de y hacia los actores, reúnen los requisitos en los límites del sistema, suelen ser abstracciones de ventanas, formularios, interfaces de impresoras, sensores, terminales, y describan lo que se obtiene con la interacción, no el proceso físico de cómo se ejecuta.

**Clase de Control:** Representan coordinación, secuencia, transacciones, y control de otros objetos, Permiten manejar los eventos del caso de uso, transacciones; usualmente encapsulan el control de un caso de uso. Las clases control además encapsulan cálculos complejos, lógica de negocio. Se podría pensar en una clase control como que “ejecuta” o “implementa” el caso de uso. Estas clases se usan durante el “análisis” para representar comportamiento complejo no asociado a un objeto entidad específico.

**Clase de Entidad:** Modelan información que posee una larga vida, modelan la información y el comportamiento asociado a algún fenómeno o concepto, como una persona, un objeto del mundo real, o un suceso del mundo real, derivan normalmente de una clase de entidad del negocio, puede tener comportamiento complejo, aísla los cambios en la información que representa.

**I**

**Interfaz:** Una colección de operaciones que se usan para especificar el servicio de una clase o de un componente. Un juego nombrado de operaciones que caracterizan la conducta de un elemento. La Interfaz hombre-máquina es un canal comunicativo entre el usuario y el ordenador.

**Isoenzimas:** Proteínas producida por un gen individual o por un grupo de ellos, que tienen una estructura química diferente pero una función enzimática similar.

**N**

**Normalizar imagen:** Ajusta la posición de referencia de las bandas respecto al sistema de referencia.

**P**

**PCR:** (Reacción en cadena de la polimerasa, en inglés The Polymerase Chain Reaction), no es más que la producción de grandes cantidades a partir de un fragmento específico de ADN, mediante la amplificación selectiva de una simple molécula de ADN muchísimas veces y en unas pocas horas.

**R**

**RFLP:** (Fragmentos de restricción de longitud polimórfica). Son regiones de gran variabilidad que difieren en cada persona (salvo entre gemelos).

**T**

**Taxonomía:** Es una forma clara y ordenada en la cual se ordenan todos los organismos vivos. Se forman de una colección de grupos llamados taxones subdivididos en distintos rangos o categorías taxonómicas.

**U**

**UML:** "Unified Modeling Language" Lenguaje gráfico que brinda un vocabulario y reglas para especificar, construir, visualizar y documentar los artefactos de un sistema utilizando el enfoque orientado a objetos.

**USB:** (en inglés Universal Serial Bus) es un Protocolo de transferencia de datos para conectar dispositivos a una computadora. Un sistema USB tiene un diseño asimétrico, que consiste en un solo servidor y múltiples dispositivos conectados en una estructura de árbol utilizando dispositivos hub especiales.