Taller de Desarrollo de Proyectos I (75.45)

FIUBA

Trabajo Práctico

Sistema de Diagnóstico Médico

|  |  |
| --- | --- |
| **Alumno** | **Padrón** |
| Adriano Pacini | 85945 |
| Alejandro Lamprópulos | 85838 |
| Ariel Scarpinelli | 85927 |
| Cristian Santilli | 86097 |
| Sebastian Durandeu | 85715 |
| Mariana Gambande | 85964 |

1er Cuat. 2009

**Contenido**

[Objetivo 4](#_Toc239093717)

[Alcance 4](#_Toc239093718)

[Cronograma de Trabajo Inicial 5](#_Toc239093719)

[Cronograma de Trabajo Real 6](#_Toc239093720)

[Backlog de Tareas 7](#_Toc239093721)

[Introducción al Problema 8](#_Toc239093722)

[Hipótesis 17](#_Toc239093723)

[Diagrama de Procesos de un Encuentro 18](#_Toc239093724)

[Análisis Estructurado 19](#_Toc239093725)

[Diagrama de Entidad-Relación 20](#_Toc239093726)

[Diagrama de Entidad-Relación 20](#_Toc239093727)

[Lista de eventos 21](#_Toc239093728)

[Lista de Estímulos y Respuestas 21](#_Toc239093729)

[Diccionario de Datos 22](#_Toc239093730)

[Entidades Externas 23](#_Toc239093731)

[Diagrama de Flujo de Datos 24](#_Toc239093732)

[Modelo 4+1 25](#_Toc239093733)

[Diagrama de Casos de Uso 26](#_Toc239093734)

[Vista Lógica - Diagrama de Clases (Modelo de Dominio) 27](#_Toc239093735)

[Vista Lógica - Diagrama de Secuencia 28](#_Toc239093736)

[Vista Lógica - Diagrama de Paquetes 29](#_Toc239093737)

[Vista de Componentes 30](#_Toc239093738)

[Vista de Procesos 31](#_Toc239093739)

[Vista de Despliegue 32](#_Toc239093740)

[Anexo A: Minutas de Reunión 33](#_Toc239093741)

[Minuta Reunión 2009-06-06 33](#_Toc239093742)

[Anexo B: Esquemas de Pantallas para Diseño de Interfaz 34](#_Toc239093743)

# Objetivo

Desarrollar un sistema experto que permita ayudar al médico en el diagnostico de un paciente. Este sistema será capaz de almacenar información de pacientes y sus encuentros, y utilizar la misma como base estadística para generar mejores diagnósticos. El sistema requerirá como datos los síntomas del paciente, los resultados de análisis clínicos y otros hechos relevantes, y, utilizando estos y una base de conocimiento, deberá devolver un conjunto de posibles diagnósticos cada uno acompañado de su probabilidad de ocurrencia.

# Alcance

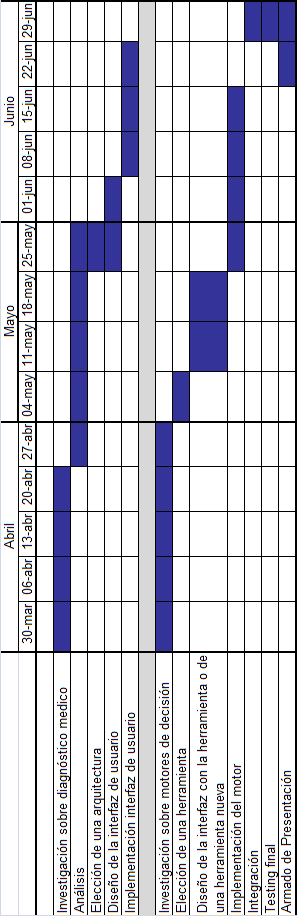
Esta dentro del alcance de este sistema:

* Elaborar una lista ordenada de diagnósticos (junto con sus correspondientes probabilidades) a partir de los síntomas, características e historia clínica de un paciente.
* Aprender de los casos reales de los pacientes para mejorar la exactitud de los diagnósticos.
* Mantener la información sobre los pacientes (características, historia clínica, etc.).
* Permitir el ingreso de casos predefinidos para tener una base estadística.

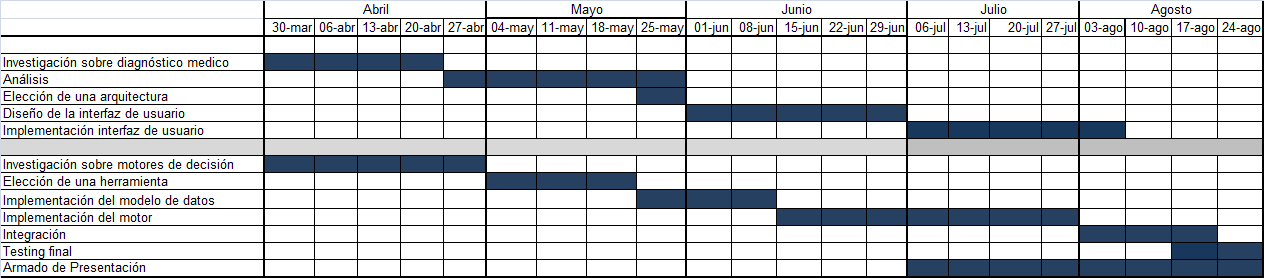
Esta fuera del alcance de este sistema:

* Realizar inferencias médicas sobre los datos más allá del cálculo de probabilidades de los diagnósticos.

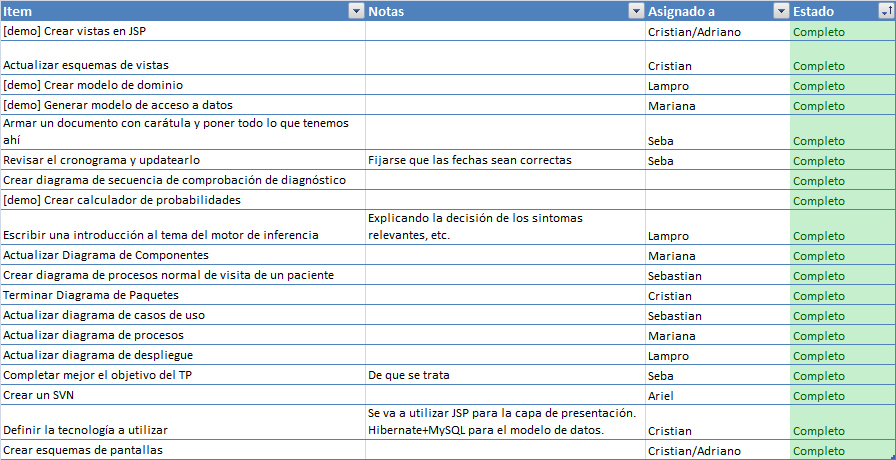
# Cronograma de Trabajo Inicial



# Cronograma de Trabajo Real



## Backlog de Tareas



# Introducción al Problema

El núcleo de la aplicación a desarrollar en este trabajo práctico consiste en un sistema experto. Un sistema experto puede deﬁnirse como un sistema informático (hardware y software) que simula a los expertos humanos en un área de especialización dada. Principalmente existen dos tipos de sistemas expertos:

* Basados expertos basados en reglas
* Basados expertos probabilísticos (también llamados basados en redes bayesianas)

El núcleo de los sistemas expertos basados en reglas es el conjunto de reglas (valga la redundancia) que describen las relaciones entre los objetos (variables). Por otro lado, en los sistemas expertos probabilísticos las relaciones entre las variables se describen mediante su función de probabilidad conjunta. Por ello, la función de probabilidad conjunta forma parte de lo que se llama conocimiento.

Nuestro sistema experto probabilístico debe ser capaz de calcular la probabilidad que tiene un paciente de tener una enfermedad dado que presenta ciertos síntomas. Lo cual puede verse matemáticamente como la probabilidad de que se presente la enfermedad *E* =  dados los síntomas . O sea, si tomamos  y  como variables, la primera corresponde a una determinada enfermedad dependiendo de su valor (supongamos m valores posibles, dado que se consideran m enfermedades), y la segunda corresponde a síntomas, habiendo n variables (cada una representa a un síntoma) y sus valores posibles son dicotómicos (1 el síntoma esta presente, 0 el síntoma no está presente).

Por lo tanto, lo que se quiere averiguar en definitiva es:



Que podemos expresarlo para más simpleza como:



Esto significa, la probabilidad condicional de que *E* valga  (se presente la enfermedad ) dado que están presentes los síntomas  con valor  (por ejemplo, dado que  = 1,  = 0..  = 1, lo que representa que está presente el síntoma 1, no lo está el 2, y está presente el síntoma n).

Los sistemas expertos memorizan información en una base de datos. Esta debe proveernos de la información necesaria para obtener el conocimiento que nos permita calcular la probabilidad buscada. Sin embargo, la base de datos puede contener información histórica de pacientes que presentaron ciertos síntomas en el contexto de una enfermedad conocida. O sea, la base de datos puede almacenar la cantidad de pacientes (frecuencia) que presentaron cierta combinación de síntomas en el contexto de una enfermedad, y estas frecuencias podemos utilizarlas para obtener probabilidades.

Mediante estas probabilidades, somos capaces de calcular la probabilidad de que un paciente tenga cierta enfermedad mediante la aplicación del teorema de Bayes. De esta forma, quedaría:



Dado que es un mismo valor (constante) para todas las enfermedades, puede simplificarse a:

Estas probabilidades, son las que nos permitirán obtener los resultados buscados, y son las que conforman lo que se llama **base de conocimiento**. Por lo tanto, lo que debemos lograr es transformar la información que nos provee la base de datos a conocimiento conocido como la base de conocimiento.

A continuación explicaremos lo que implica la base de conocimiento con más detalle y veremos las diferentes formas de representarla.

La Base de Conocimiento

La base de conocimiento de un sistema experto basado en reglas consta del conjunto de objetos (variables) y del conjunto de reglas. La base de conocimiento de un sistema experto probabilístico, en cambio, consiste en un conjunto de variables, {X1, . . . , Xn}, y una función de probabilidad conjunta definida sobre ellas, p(x1, . . . , xn). Por ello, para construir la base de conocimiento de un sistema experto probabilístico, se necesita definir la función de probabilidad conjunta de las variables.

El modelo mas general posible se basa en especificar directamente la función de probabilidad conjunta; es decir, asignar un valor numérico (parámetro) a cada una de las posibles combinaciones de valores de las variables. Desgraciadamente, la especificación directa de la función de probabilidad conjunta implica un gran número de parámetros, es por esto que se estudian diferentes modelos para representar a la base de modo de conseguir manejar una menor cantidad de parámetros intentando representar la realidad de la mejor forma posible.

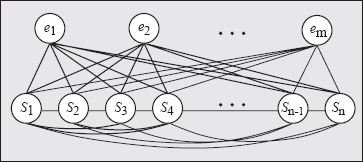
Hay cuatro modelos que podemos tener en consideración:

* Modelo de síntomas dependientes
* Modelo de síntomas independientes
* Modelo de síntomas relevantes independientes
* Modelo de síntomas relevantes dependientes

Modelo de Síntomas Dependientes

En este modelo se supone que los síntomas son dependientes pero que las enfermedades son independientes entre si, dados los síntomas.

En la figura puede observarse el modelo en forma grafica.



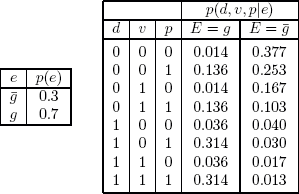
Utilizando esta forma de representar la realidad, el cálculo de la probabilidad puede expresarse como habíamos visto anteriormente:



Donde  es la probabilidad marginal de tener la enfermedad  a priori (que puede calcularse simplemente como número de casos de esa enfermedad sobre número de casos totales), y es la probabilidad de presentar una combinación de síntomas dado que se tiene la enfermedad  (que también es calculable como la cantidad de casos que presentaron esa combinación de síntomas, teniendo la enfermedad , sobre el total de los casos que la presentaron).

Por ejemplo, supongamos que tenemos un una enfermedad a la que llamaremos g. Y en nuestra base de datos contamos solamente con tres síntomas posibles, a los que le daremos los nombres d, v, p. La base de conocimiento estaría dada por pares de tablas (una para cada enfermedad) de la siguiente forma:



Esto no es más que las distribuciones de probabilidad de las distintas variables, y al ser dependientes los síntomas, las variables que los representan deben presentar todas las distintas combinaciones con sus correspondientes probabilidades.

Véase que en la tabla se indica también la probabilidad de no tener la enfermedad y de las distintas combinaciones de síntomas dado que no se presenta la enfermedad. Esta es información que puede obtenerse de la base de datos, y se necesita debido a que al quitar el factor de normalización constante para todos los cálculos, se necesitará calcular tanto  como  para poder normalizarlos a probabilidad 1.

Para este ejemplo en particular, si tomamos que el paciente presenta los tres síntomas, vemos que:

Por lo tanto, normalizando nos queda que 0.2198+0.0039 = 0.2237, y entonces:

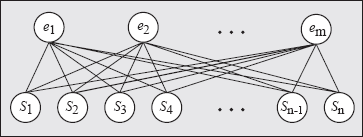
= 0.2198 / 0.2237 = 0.98

O sea, que la probabilidad de tener la enfermedad dado que el paciente presenta los tres síntomas es de 0.98.

El principal problema de este modelo es que requiere un número muy alto de parámetros. Claramente, especificar las probabilidades para todas esas combinaciones es muy difícil y se hace imposible al crecer los números de enfermedades y síntomas. Por ejemplo, con 100 enfermedades y 200 síntomas (que no es una situación irreal), el numero de parámetros necesarios es mayor a , tan grande que sería completamente inmanejable. Es por esto que se presentan otras alternativas.

Modelo de Síntomas Independientes

Debido a la imposibilidad de trabajar con el modelo anterior en muchos casos prácticos, resulta necesario proceder a la simplificación del mismo. Una posibilidad consiste en suponer que, para una enfermedad dada, los síntomas son condicionalmente independientes entre si. Este es el modelo de síntomas independientes, que puede observarse en la siguiente figura.



Como podemos ver, las enfermedades siguen siendo independientes entre si, pero ahora también los síntomas lo son, y además, las enfermedades dependen de todos los síntomas del sistema.

Nuestro cálculo ahora se puede expresar de la forma:

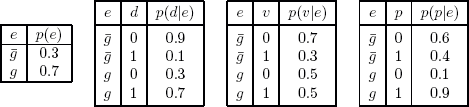




Siendo  la probabilidad condicional de que se presente o no el síntoma  dado que se tiene la enfermedad .

Véase que gracias a la hipótesis de independencia, pudimos transformar  en una productoria . Esto nos permite que nuestra base de conocimiento sea de proporciones más pequeñas, quedando de la forma (para el ejemplo anterior):

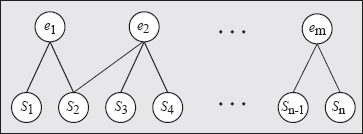


Para realizar los cálculos, se procede de la misma manera que antes, calculando la probabilidad de tener la enfermedad y de no tenerla, dada la combinación de síntomas, pero ahora aplicando la nueva fórmula para calcularla.

Aunque la hipótesis de independencia da lugar a una gran reducción del número de parámetros, el mismo es todavía muy alto para ser práctico. Por lo tanto se necesita simplificar el modelo aun más.

Modelo de Síntomas Relevantes Independientes

Se puede conseguir una reducción aun mayor del número de parámetros suponiendo que cada enfermedad tiene un número reducido de síntomas relevantes, lo cual es bastante coherente con la realidad. En consecuencia, para cada valor de la enfermedad *E* se seleccionan algunos síntomas relevantes  (relativamente pocos frente al total de síntomas) y los restantes síntomas se suponen independientes para ese valor de *E*. En la siguiente figura puede observarse como es el modelo de manera gráfica:



Suponiendo entonces que los primeros *r* síntomas son relevantes para la enfermedad  y que los restantes son irrelevantes, nuestro calculo de probabilidad quedaría de la siguiente forma:







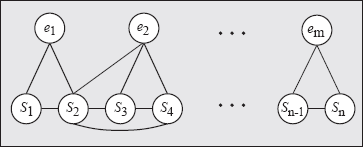
Véase que se tomó  como , esto es porque para este modelo las probabilidades que llamamos , para cada valor posible de *E* que tiene al menos un síntoma irrelevante, es idéntica para todos los síntomas irrelevantes de , y es de esta forma que se consigue una simplificación del modelo.

De esta manera, el nuevo modelo reduce mucho la cantidad de parámetros, con la condición de que se conozcan cuales son los síntomas relevantes de cada una de las enfermedades, conocimiento que puede obtenerse de la ciencia médica.

Modelo de Síntomas Relevantes Dependientes

Aunque el modelo anterior reduce el número de parámetros considerablemente, por desgracia, es poco realista, ya que los síntomas asociados a ciertas enfermedades suelen producirse en grupos o síndromes. Por ello, puede ser poco razonable suponer que los síntomas relevantes son independientes. El modelo de síntomas relevantes dependientes evita este inconveniente, cambiando solamente la consideración de que los síntomas relevantes en vez de se independientes, sean dependientes, dada la correspondiente enfermedad. De esta forma, se supone que solo los síntomas irrelevantes son independientes pero los relevantes pueden ser dependientes.

La siguiente figura muestra el modelo gráficamente:



Suponiendo entonces que los primeros *r* son los síntomas relevantes para la enfermedad , y que el resto son irrelevantes, la probabilidad de que se presente una enfermedad, dada una combinación de síntomas, puede expresarse de la siguiente forma:





Siendo nuevamente las probabilidades , para cada valor posible de *E* que tenga al menos un síntoma irrelevante, idénticas para todos los síntomas irrelevantes para . Y además, siendo las probabilidades condicionales, para todos los posibles valores de la enfermedad *E* y sus síntomas relevantes .

De esta forma, la base de conocimiento queda limitada a almacenar (para cada enfermedad) las probabilidades marginales de tener o no la enfermedad, y las probabilidades de presentar las distintas combinaciones de síntomas relevantes dado que se tiene o no la enfermedad.

Gracias a la utilización de estos modelos, vemos simplificada la cantidad de parámetros representando la realidad de la manera más fiel posible. No obstante, observamos que no solamente los síntomas pueden ser evidencia para detectar una enfermedad, sino que cuando un paciente acude al médico, este puede tener en cuenta otros parámetros para la determinación de la misma. Es por esto que se elige representar a todos los posibles factores que pueden determinar una enfermedad, ya sean síntomas que presenta el paciente, parámetros clínicos tomados por el médico al paciente, características físicas de este ultimo (ya sea la edad, si fuma o no, etc.), o hasta enfermedades que tuvo el paciente en el pasado. Y esto último nos da la posibilidad de expandir nuestro modelo teniendo en cuenta que las enfermedades pueden ser dependientes entre sí, y tomándolas como un parámetro mas de evidencia.

Un detalle que puede tenerse en cuenta es que hay parámetros como la fiebre, la edad, entre otras, que no pueden medirse con un valor dicotómico sino que son parámetros continuos, lo cual debe simplificarse aplicando rangos sobre los mismos. O sea, en el caso de la fiebre se pueden definir cuatro parámetros de evidencia diferentes ya sea alta, media, baja, no fiebre, y encasillando los valores de temperatura del paciente dentro de alguno de estos.

Resumiendo, se eligió aplicar el último de los modelos (modelo de síntomas dependientes) junto a algunos parámetros extra recién descriptos (como agregar parámetros clínicos, enfermedades, y características como evidencia) para la implementación de nuestra base de conocimiento. Con esto el motor de inferencia se encargará de realizar los cálculos necesarios para obtener las probabilidades de que un paciente tenga cada una de las enfermedades que el sistema conozca. El cálculo deberá realizarse para todas ellas, y se devolverán los resultados de las enfermedades más probables, siempre teniendo en cuenta toda la evidencia relevante que el médico pueda tomar sobre el paciente.

# Hipótesis

Para el cálculo de probabilidades del motor de inferencia se tomarán las siguientes hipótesis:

* Las enfermedades entre sí son independientes
* Los síntomas relevantes son dependientes
* Los síntomas no relevantes son independientes

También se tienen las siguientes hipótesis sobre el tratamiento de la información histórica:

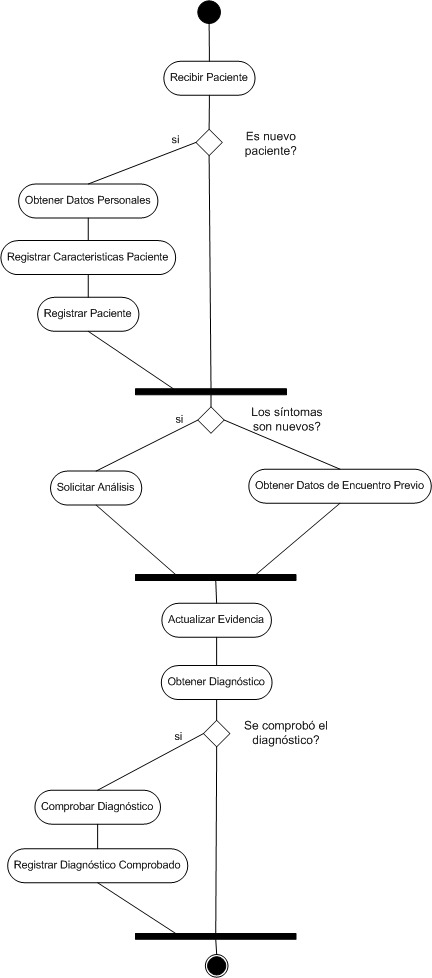
* Existen estadísticas pre-existentes (históricas) sobre la cantidad de casos de una enfermedad dada. Cada caso de enfermedad histórico también tiene definido los síntomas que se observaron en el paciente.
* Los diagnósticos que se comprueban empíricamente son tomados como información histórica para el cálculo de nuevas probabilidades.

Además el sistema se comportará de la siguiente forma:

* El sistema no recalculará diagnósticos ya realizados al incorporar más información para los mismos (aprendizaje), el médico será el encargado de volver a solicitar los diagnósticos si así lo requiere.
* El sistema no permitirá mantener varios valores del mismo parámetro dentro de un encuentro. Si bien el médico puede editar un encuentro, al cambiar un valor, el original se pierde.

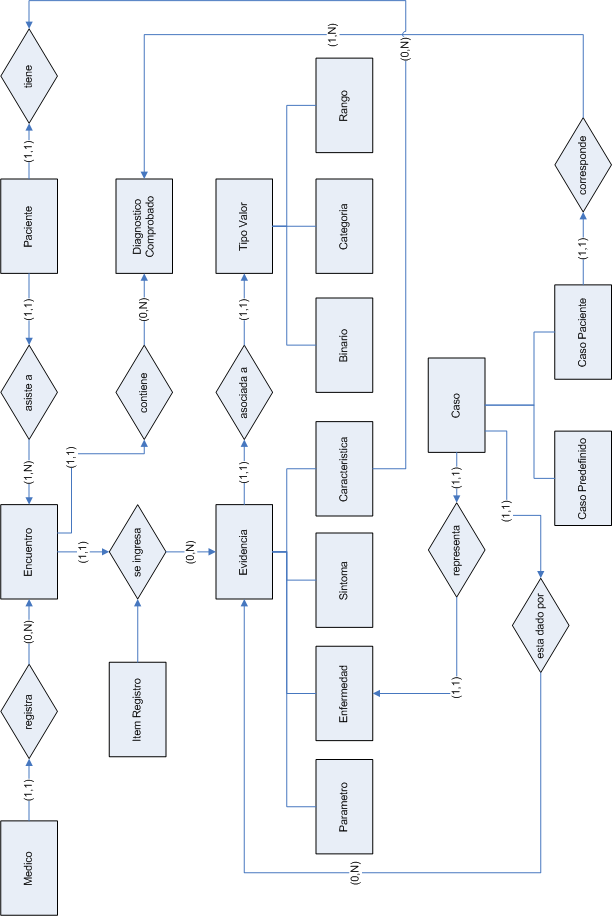
# Diagrama de Procesos de un Encuentro

El sistema de desarrollar será utilizado por un médico al momento de recibir al paciente. A esta situación de la denomina “encuentro”. En este proceso ocurren las siguiente actividades:



|  |
| --- |
| Análisis  Estructurado |

## Diagrama de Entidad-Relación



Lista de eventos

1. Medico registra encuentro.
2. Medico registra diagnostico comprobado.
3. Medico mantiene paciente.
4. Administrador registra casos predefinidos.
5. Administrador registra enfermedad.
6. Administrador registra síntoma.
7. Administrador registra parámetro clínico.
8. Administrador registra característica de pacientes.
9. Administrador registra médicos usuarios.

## **Lista de Estímulos y Respuestas**

1) Medico 🡪 datos encuentro   
    Medico 🡨 diagnostico  
               < encuentros >  
  
2) Medico 🡪 diagnostico comprobado  
               < diagnósticos >  
  
3) Medico 🡪 datos paciente  
               < pacientes >  
  
4) Administrador 🡪 caso predefinido  
               < casos predefinidos >  
  
5) Administrador 🡪 enfermedad  
               < enfermedades >  
  
6) Administrador 🡪 síntoma  
               < síntomas >  
  
7) Administrador 🡪 característica  
               < características >  
  
8) Administrador 🡪 parámetro clínico  
               < parámetros clínicos >  
  
9) Administrador 🡪 medico usuario  
               < médicos >

## Diccionario de Datos

* **Médico Usuario:** Nombre + Matrícula + Especialidad  
  Descripción: “Datos relativos al médico como usuario del sistema.”
* **Datos Paciente:** Nombre + DNI + Fecha Nacimiento + Dirección + Teléfono + {Característica}
* **Datos Encuentro:** {Evidencia} + Fecha + Identificador Paciente + Identificador Médico  
  Descripción: “Síntomas y parámetros clínicos del paciente observados por el médico, necesarios para obtener un diagnóstico.”
* **Evidencia:** “El conjunto de síntomas y parámetros clínicos observables o detectables en el paciente, que el médico ingresa al sistema durante un encuentro, para obtener un diagnóstico.”
* **Característica:** Nombre + Tipo Valor  
  Descripción: “Dato inherente a las cualidades físicas del paciente independientemente de si padece o no enfermedades. Ejemplo: Peso, Altura, Presión normal, etc.”
* **Diagnóstico:** {Enfermedad, Probabilidad, Tratamiento Recomendado} + Identificador Paciente  
  Descripción: “Una lista ordenada de posibles enfermedades que padece el paciente.”
* **Diagnóstico Comprobado:** Identificador Medico + Identificador Paciente + Identificador Encuentro + Duración Temporal + Enfermedad  
  Descripción: “Confirmación de la enfermedad que empíricamente se comprobó que tenía el paciente al cual se le generó un diagnóstico.”
* **Caso Predefinido:** Enfermedad + {Síntoma}  
  Descripción: “Caso de enfermedad que se ingresa al sistema como información histórica para el cálculo de probabilidades”
* **Enfermedad:** Nombre Enfermedad + Tratamiento Recomendado + Cronicidad + Unicidad + Duración Temporal  
  Descripción: “Una alteración de la salud del paciente, con una probabilidad de ocurrencia asociada.”
* **Síntoma:** Nombre Síntoma + Tipo Valor  
  Descripción: “Cualquier manifestación observable o detectable, no cuantificable, que sea relevante para la determinar el diagnóstico del paciente.”
* **Parámetro Clínico:** Nombre Parámetro + Unidad + Tipo Valor

Descripción: “Cualquier manifestación cuantificable del estado de salud del paciente que sea relevante para determinar el diagnóstico del paciente”

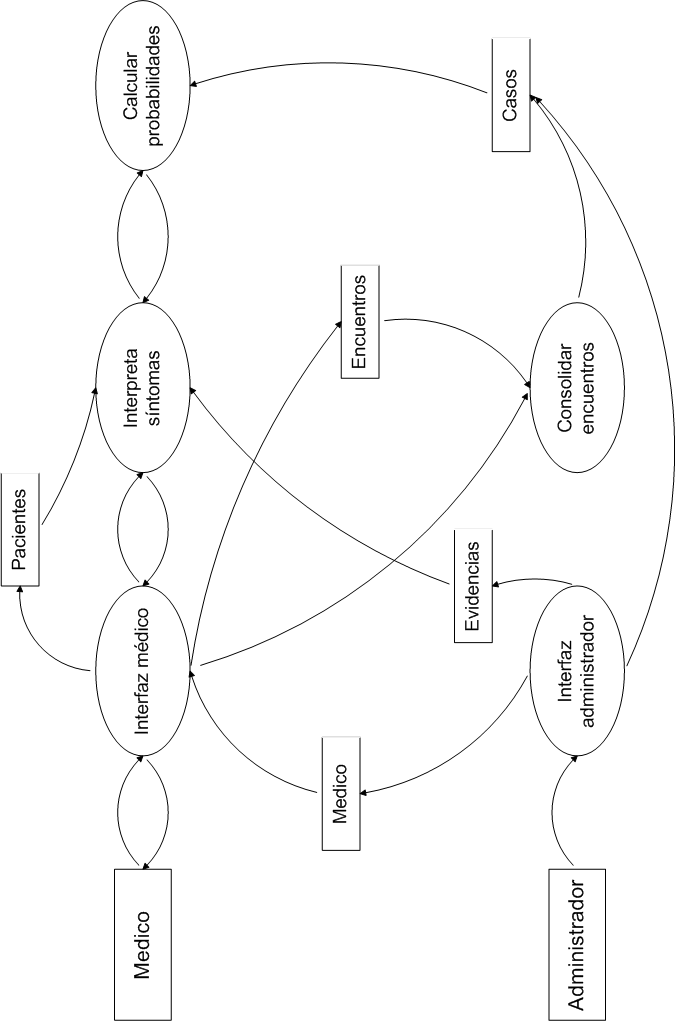
* **Tipo Valor:** [Binario | {Categorías} | Rango Válido]  
  Descripción: “Forma en la que se cuantifica la evidencia del paciente (ver descripción evidencia).”

**Nota:** Las llaves (“{}”) simbolizan la ocurrencia repetida del componente de un dato.

## Entidades Externas

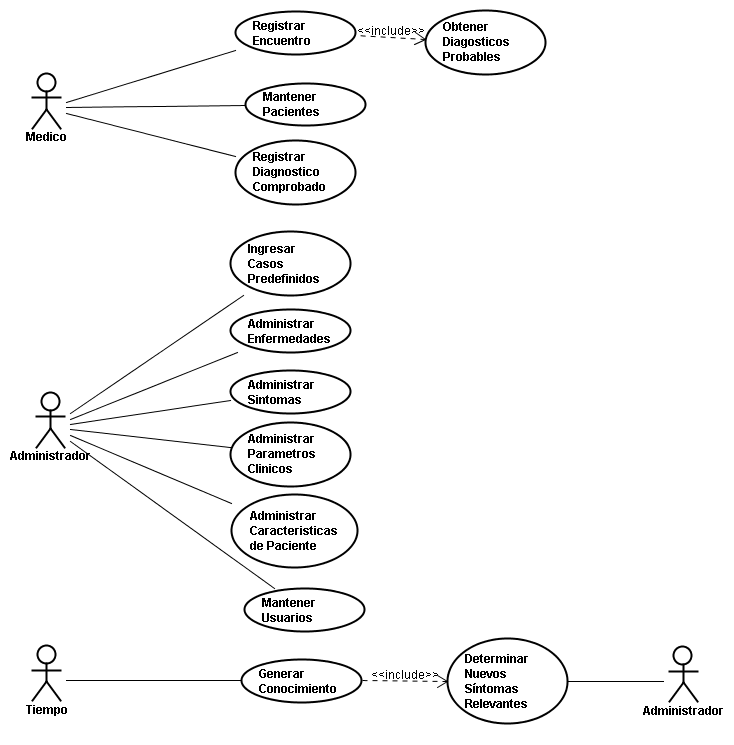
* **Médico:** Es el usuario principal del sistema. Encargado de ingresar los datos del encuentro al sistema para obtener un diagnóstico.
* **Administrador:** Es el encargado de mantener los datos necesarios para que el sistema funcione correctamente. Define las enfermedades, parámetros clínicos y síntomas soportados por el sistema. Además es el encargado de registrar información histórica de casos predefinidos.

## Diagrama de Flujo de Datos

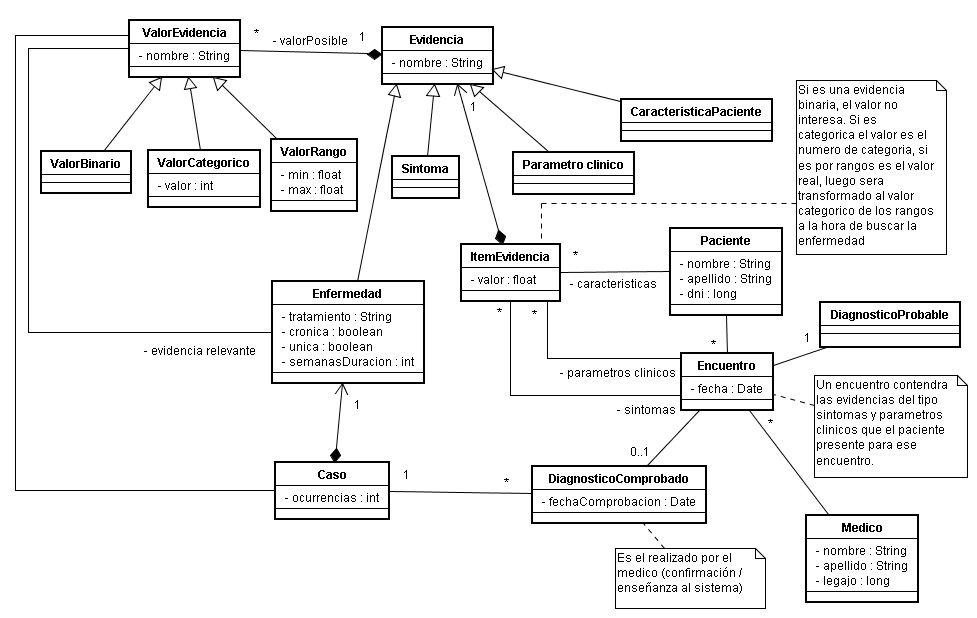


|  |
| --- |
| Modelo  4+1 |

## Diagrama de Casos de Uso

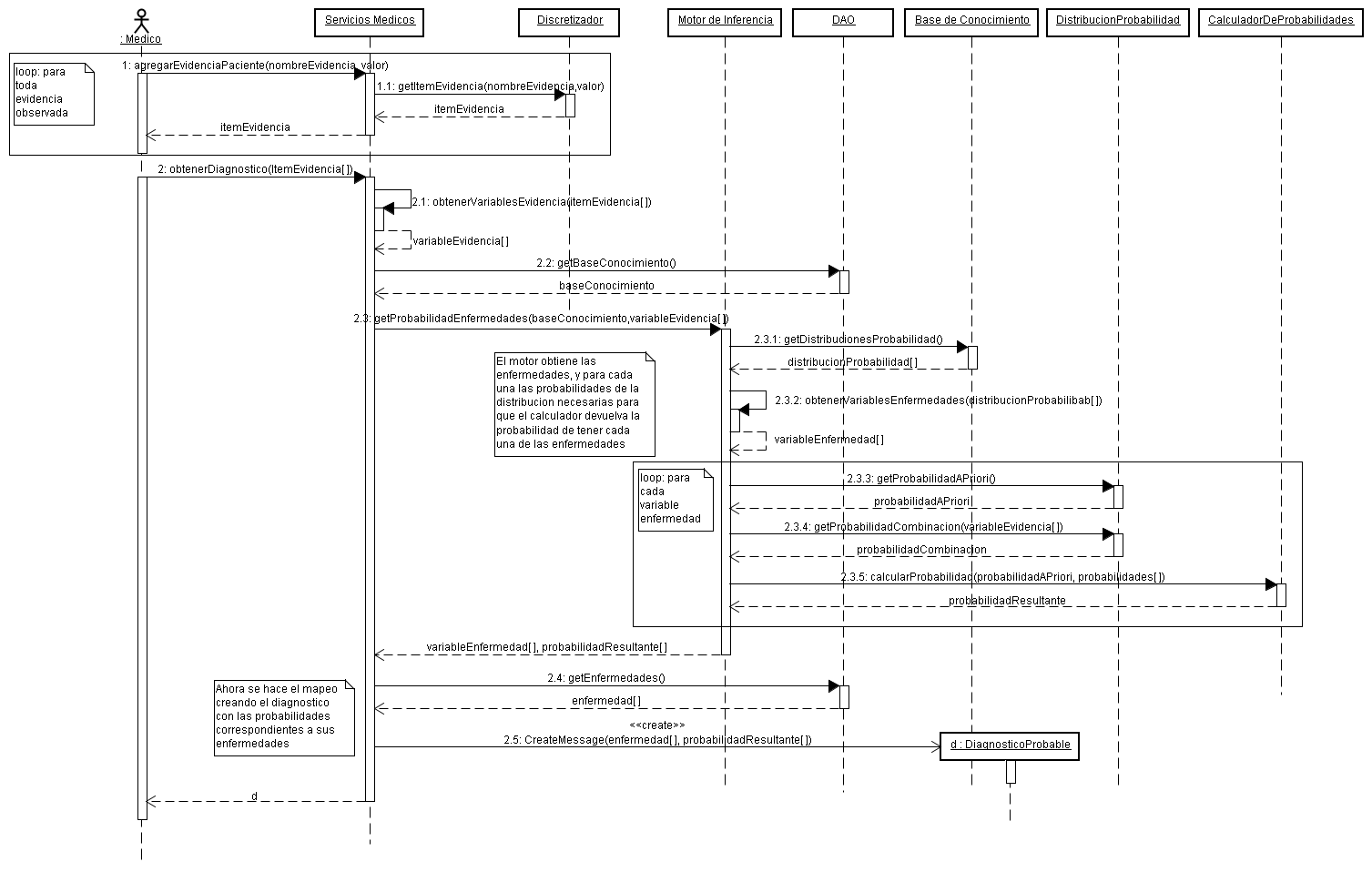


## Vista Lógica - Diagrama de Clases (Modelo de Dominio)

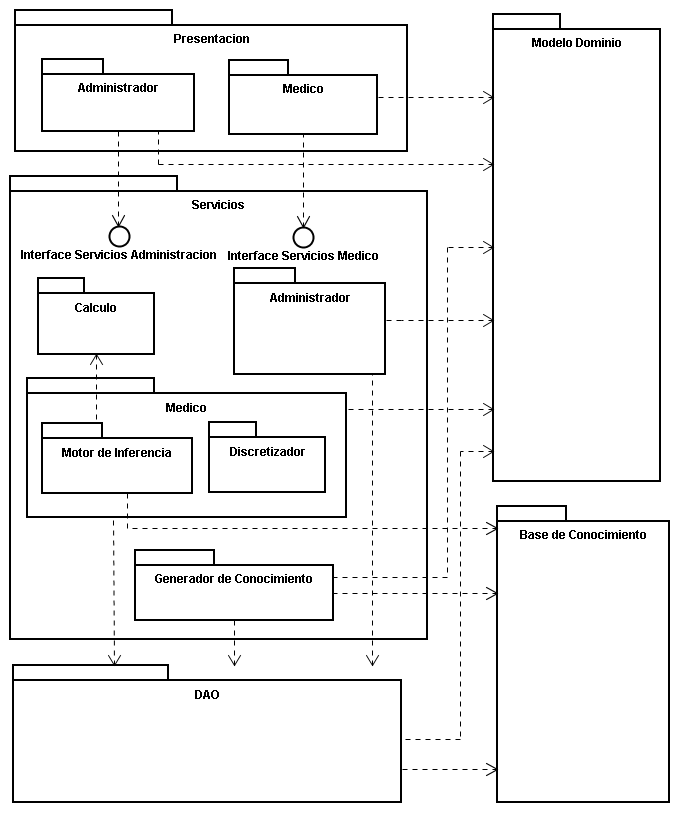


## Vista Lógica - Diagrama de Secuencia

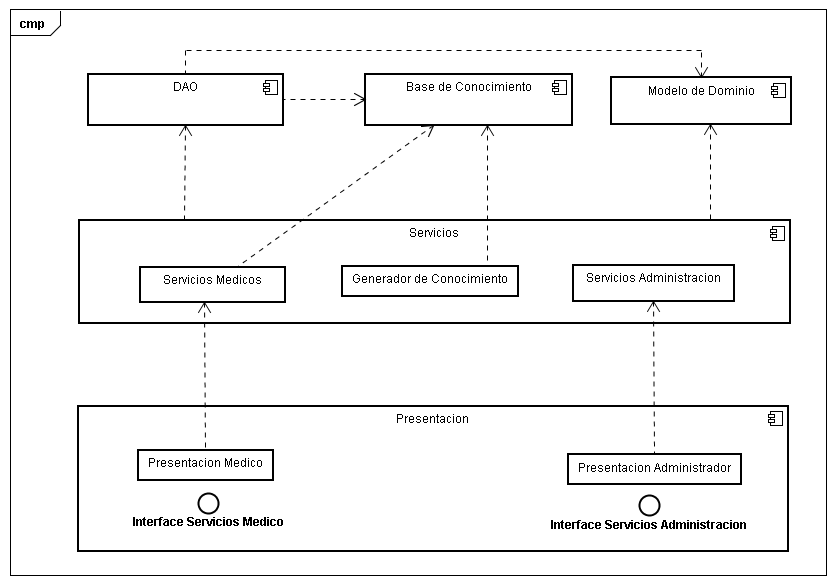
**Escenario**: Obtener un diagnóstico en base a evidencia.



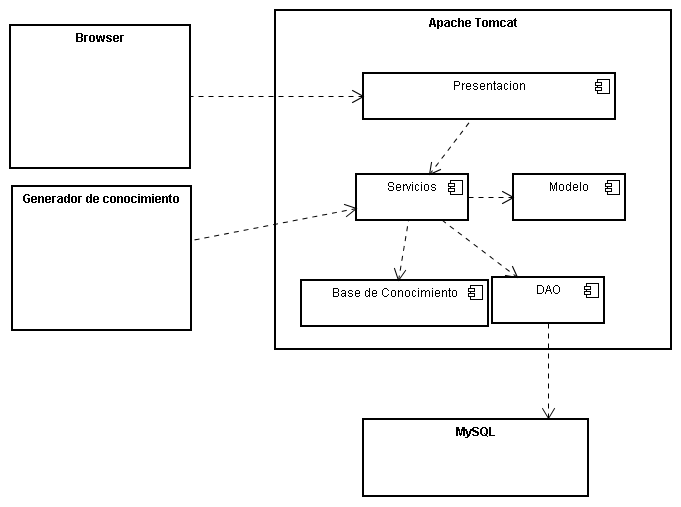
## Vista Lógica - Diagrama de Paquetes



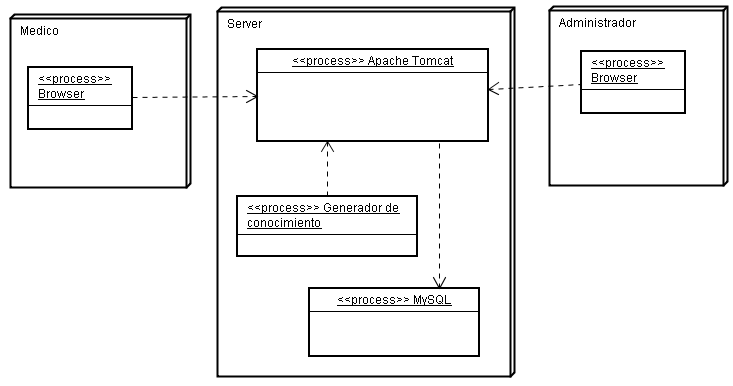
## Vista de Componentes



## Vista de Procesos



## Vista de Despliegue



Anexo A: Minutas de Reunión

## Minuta Reunión 2009-06-06

**Decisiones del Modelo**

- Necesitamos relacionar distintos encuentros físicos entre el paciente y el médico.

- El médico va a poder editar encuentros viejos y usarlos para obtener un diagnóstico. El encuentro va a quedar guardado con la última fecha en la que se editó. El médico es responsable si modifica valores viejos que eran relevantes.

- Puede haber encuentros que no tengan diagnóstico comprobado.

**Decisiones del Motor**

- Para normalizar el valor de probabilidad obtenido calculamos la probabilidad de tener la enfermedad y la probabilidad, de no tenerla.

- Vamos a tener un cache de probabilidades que se va a actualizar cada cierto tiempo. Esto esta modelado por el Servicio de Realimentación y el proceso de Realimentación.

- El administrador va a definir cuales síntomas son relevantes.

- El Pj (constante para todos los síntomas irrelevantes) al normalizar se pierde.

- Todo síntoma que el médico no ingresa, se considera como que el paciente no lo tiene.

**Cosas Relevantes del Libro**

- Buena figura para diseñar la base de datos --> FIGURA 3.1

- Fórmula --> 3.29

- Probabilidades Requeridas para el modelo --> TABLA 3.14

**Decisiones de la Demo**

- Login en cada presentación: Administración y Médico

Anexo B: Esquemas de  
Pantallas para Diseño de Interfaz