

**ESTRATEGIA CO - EVOLUTIVA PARA LA IDENTIFICACIÓN DE MODELOS DE
AUTÓMATA CELULAR EN TRAYECTORIAS DE SIMULACIÓN DE PLEGAMIENTO DE
PROTEÍNA**



ANEXOS

Adriana Victoria Gómez Buitrón

Sayra Mildreth Ocoró Rosero

Director: MSc. Néstor Milciades Díaz Mariño

Universidad del Cauca

Facultad de Ingeniería Electrónica y Telecomunicaciones

Departamento de Sistemas

Grupo de I+D en Tecnologías de la Información

Modelado y Simulación de Sistemas Complejos

Popayán, Julio de 2012

TABLA DE CONTENIDO

INDICE DE TABLAS	ii
INDICE DE FIGURAS	ii
1 ARTEFACTOS DE CRISP-DM	1
1.1 Riesgos y Contingencias.....	1
1.2 Costos y beneficios.....	2
1.3 Plan del Proyecto.....	2
2 ARTEFACTOS DE XP	5
2.1 Diagrama de Clases.....	5
2.2 Pruebas de Caja Negra.....	7
2.2.1 Formato de Pruebas de Funcionalidad Externa	7
2.2.2 Batería de Pruebas.....	11
2.2.3 Historia de Revisiones	13

INDICE DE TABLAS

Tabla 1. Tabla de riesgos y contingencias	1
Tabla 2. Tabla de recursos.....	2

INDICE DE FIGURAS

Figura 1. Cronograma de actividades.....	4
Figura 2. Diagrama de clases.....	6
Figura 3. Interfaz Aplicación co- evolutiva, Begin.....	9
Figura 4. Interfaz Aplicación co- evolutiva, Definir parámetros 1ra parte.....	10
Figura 5. Interfaz Aplicación co- evolutiva, Definir parámetros 2da parte.....	10
Figura 6. Interfaz Aplicación co- evolutiva, Set Parameters.....	11

1 ARTEFACTOS DE CRISP-DM

En esta sección del documento se muestran los artefactos sobresalientes obtenidos en las diferentes etapas de la metodología CRISP – DM. Esta documentación es útil para el entendimiento del negocio y para futuras investigaciones que tomen como referente este proyecto de investigación.

1.1 RIESGOS Y CONTINGENCIAS

A continuación se presenta en la Tabla 1 los posibles riesgos identificados en la etapa de Entendimiento del Negocio de CRISP – DM, y las actividades de contingencia, para cada uno de ellos.

Tabla 1. Tabla de riesgos y contingencias

ID	DESCRIPCIÓN	DISPARADOR	ACCION(ES) PREVENTIVA(S)	IMPACTO	PROBABILIDAD	ACTIVIDAD(ES) DE CONTINGENCIA
R1	Planificación mal realizada.	Atraso en las actividades programadas.	Monitoreo frecuente de las actividades para solucionar posibles retrasos.	Alto	Baja	Dedicar tiempo extra a las actividades retrasadas para cumplir con el tiempo estipulado.
R2	La complejidad del contexto de los datos, y el entendimiento de los mismos por parte del equipo de proyecto podrían ocasionar retrasos en la ejecución del proyecto.	El diseño de la solución toma más tiempo o es más costoso de lo estimado.	Análisis detallado del presupuesto necesario para la ejecución del proyecto.	Alto	Media	Para evitar la materialización del riesgo, se requerirán reuniones constantes con el cliente para llegar a un mejor entendimiento de los datos.
R3	Pérdida de la información del proyecto	Pérdida o daño del equipo que contiene la información del proyecto.	Realizar copias de respaldo en discos ópticos o subirlos a un servidor.	Alto	Media	Recurrir a las copias de seguridad.
R4	Mala organización de los integrantes del grupo.	Se repite el trabajo o no se realiza el trabajo por los integrantes del grupo	Verificar los roles y responsabilidades de los integrantes del equipo.	Medio	Baja	Reunión general de los integrantes del grupo para definir nuevos roles y responsabilidades.
R5	La técnica de minería de datos escogida para llevar a cabo la tarea, puede no representar el fenómeno físico de manera legible o con la calidad	Los resultados del modelo obtenido no son los esperados.	Realizar pruebas frecuentes para verificar que se	Alto	Baja	Utilizar otras técnicas, o la implementar una que permita culminar con éxito el proyecto, pero, debido a los escasos recursos disponibles,

	que se requiere para el proyecto.					tiempo y dinero, no será posible llevar a cabo la última opción
R6	Pruebas del producto mal diseñadas y ejecutadas.	La ejecución de las pruebas es correcta pero se le presentan errores al cliente.	Realizar pruebas alfa para evitar los defectos en la aplicación.	Alto	Media	Rediseñar las pruebas de la aplicación.

1.2 COSTOS Y BENEFICIOS.

A continuación se presenta la tabla de costos (Tabla 2) considerados en la etapa de entendimiento del negocio de CRISP - DM, en la que se describen los recursos necesarios para llevar a cabo el proyecto. Los costos presentados en la tabla corresponden a 2 investigadores y un asesor de proyecto, además de los equipos y la papelería necesarios.

Tabla 2. Tabla de recursos

Tabla/Justificación	Recursos		Total
	FJET - Sistemas	Estudiantes	
Personal	1.752.960	32.868.000	34.620.960
Equipo	958.333	452.000	1.410.333
Software: Python, Boa Constructor, PYRO, CAIF - PFT, Orange, Open Office.			
Linux, Open Office	0	0	0
Viajes y Salidas de campo	0	0	0
Bibliografía	0	150.000	150.000
Materiales	0	300.000	300.000
Servicios técnicos	0	300.000	300.000
Publicaciones	0	0	0
Administración	0	0	0
Comunicaciones	76.500	495.000	571.500
Otros	0	0	0
TOTAL	2.787.793	34.565.000	37.352.793

1.3 PLAN DEL PROYECTO.

La aplicación de la metodología de desarrollo XP se hará en doce iteraciones (Figura 1), en las cuales se encapsularán las actividades de cada fase de CRISP - DM:

- *Primera Iteración:* Durante la primera iteración de XP se abordará una primera parte de la fase de comprensión del negocio de CRISP - DM, donde se incluirá determinar los objetivos de negocio y evaluar la situación actual del mismo.

- *Segunda Iteración:* En este punto del proyecto, se espera concluir con la primera fase de CRISP – DM con la construcción del plan del proyecto; se abordará la fase de análisis de los datos, y la fase de procesamiento de los datos, en la que se construirán rutinas que permitan limpiar, construir y formatear datos.
- *Tercera Iteración:* Se iniciará con la fase de Modelado de CRISP – DM. En lo concerniente a la selección de la técnica de modelado, se definirá la estrategia co - evolutiva y la representación de la solución.
- *Cuarta Iteración:* Continuando con la selección de la técnica de modelado de la fase de modelado, en ésta iteración se espera definir operadores genéticos y definir función de evaluación de los individuos.
- *Quinta Iteración:* Durante ésta iteración, se finalizan las actividades de la fase de modelado definiendo cómo se evaluará la estrategia co – evolutiva. Además, en ésta iteración se iniciará con la generación del diseño de la prueba.
- *Sexta Iteración:* Se finalizará la construcción del diseño de la prueba, y se inicia con la construcción del modelo de AC.
- *Séptima Iteración:* Se finalizará con la construcción del modelo de AC y se iniciará la evaluación del mismo teniendo en cuenta las medidas descritas en [18][44].
- *Octava Iteración:* Se espera concluir la evaluación del modelo de AC e iniciar a evaluar los resultados obtenidos con la aplicación de la metodología CRISP - DM.
- *Novena Iteración:* En ésta iteración se terminará de evaluar los resultados, se retomará modelado y conforme a los resultados obtenidos al aplicar la evaluación se efectuarán las respectivas modificaciones sobre el modelo de AC.
- *Décima Iteración:* Se retorna a la fase de Evaluación. En ésta iteración la única actividad que se realizará será evaluar el Modelo de AC modificado.
- *Décima Primera Iteración:* Se concluirá con la fase de Evaluación. Las actividades a realizar durante ésta iteración serán revisar el proceso y determinar el siguiente paso.
- *Décima Segunda Iteración:* Una vez finalizadas cinco de las seis fases de CRISP – DM, se procederá a realizar el despliegue de la solución, actividad con la que se concluye la aplicación de la metodología CRISP – DM.

Documentación: en forma paralela al desarrollo de las anteriores fases, se realizará una actividad de documentación permanente.

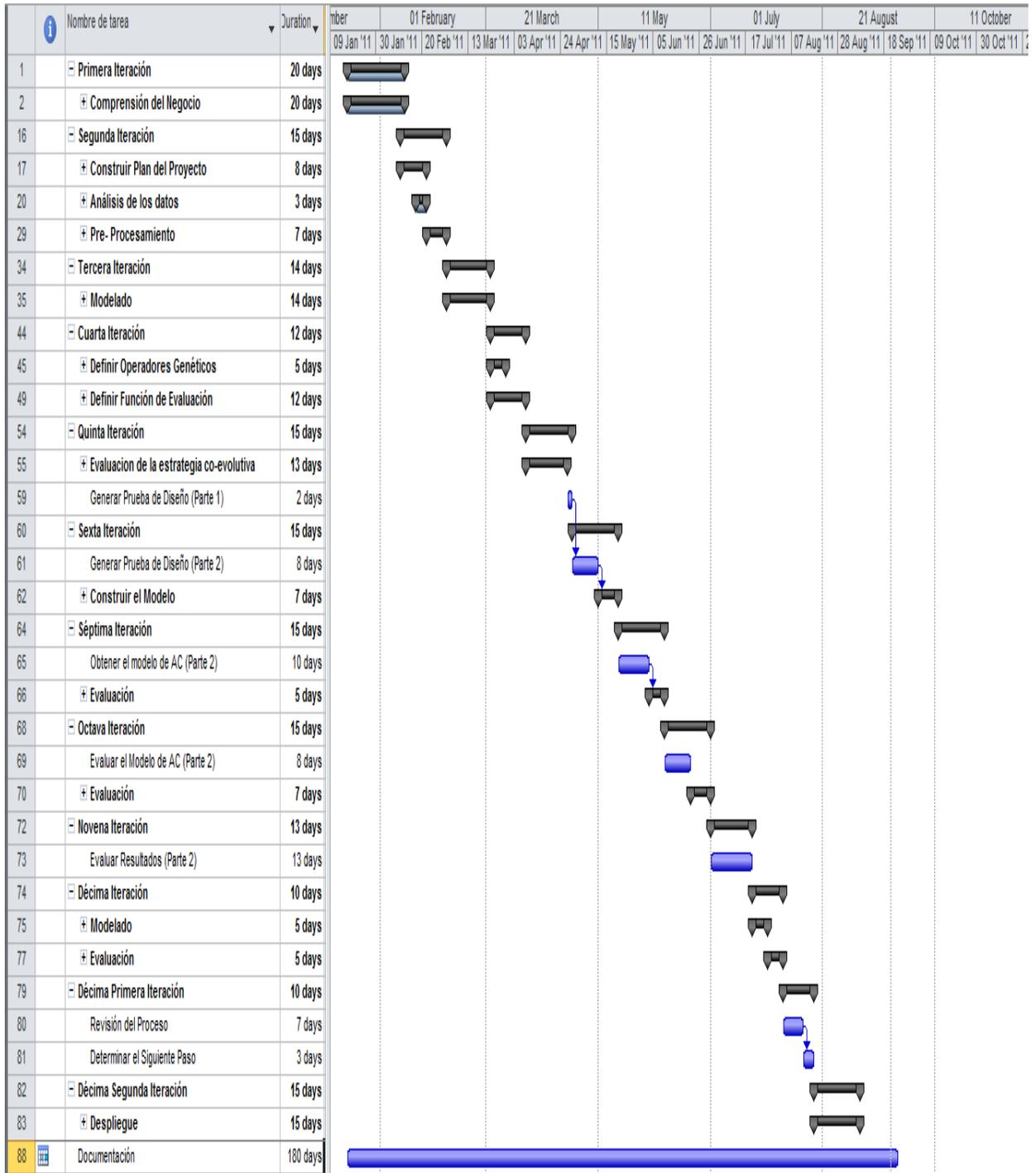


Figura 1. Cronograma de actividades.

2 ARTEFACTOS DE XP

2.1 *DIAGRAMA DE CLASES*

A continuación, se muestra el diagrama de las clases que se utilizó en la implementación de la estrategia Co – Evolutiva, es posible observar en la figura 2, el paquete COESCAIF – PFT , que contiene las clases de la aplicación implementada, además se observa el paquete del framework CAIF – PFT que es el que proporciona el soporte a la aplicación.

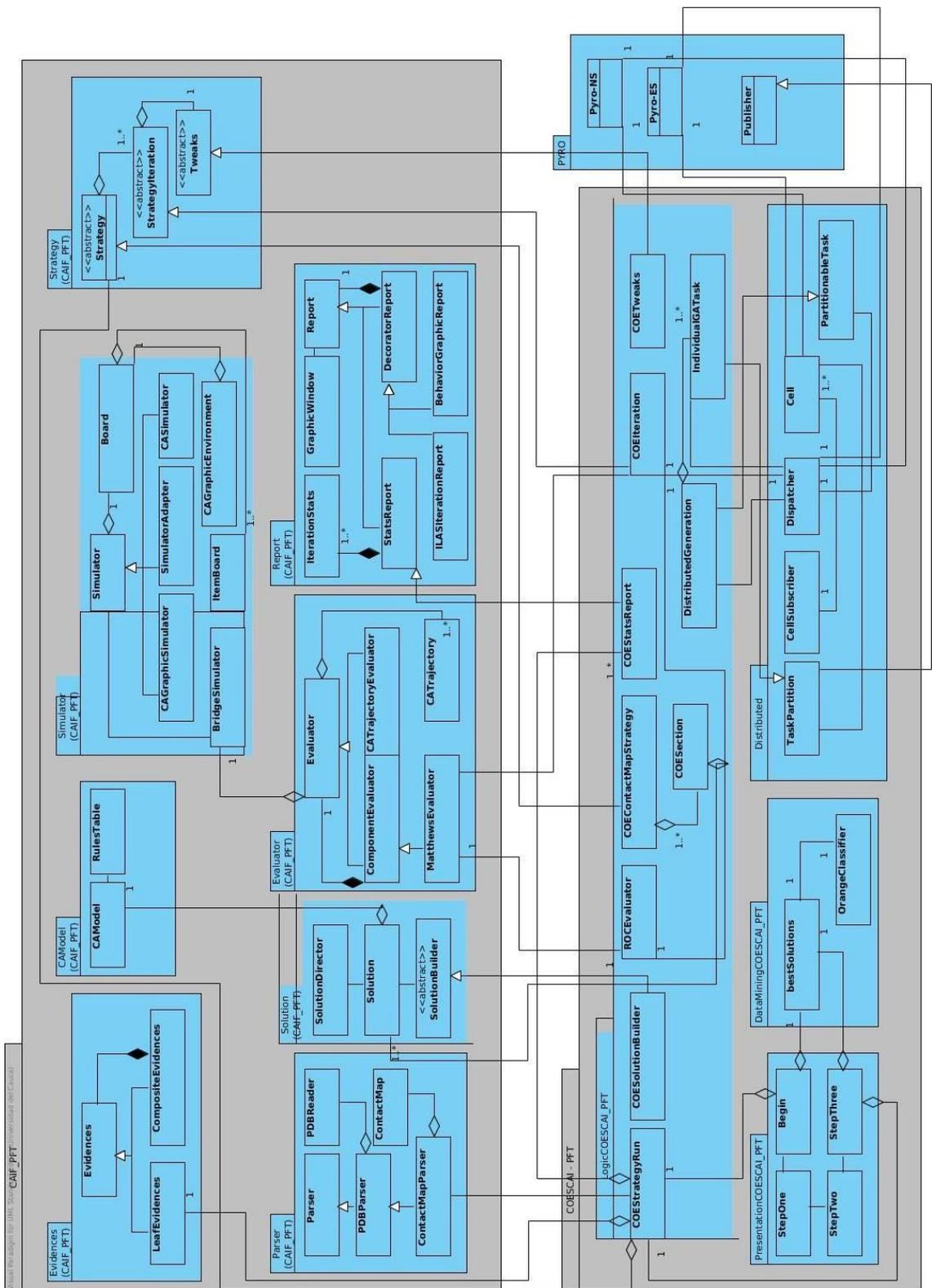


Figura 2. Diagrama de Paquetes.

2.2 PRUEBAS DE CAJA NEGRA

2.2.1 Formato de Pruebas de Funcionalidad Externa

Número de la vista	Nombre de la vista	Descripción
1	Begin (Figura 3)	En esta vista se carga o se fijan parámetros.
1.1	Definir parámetros 1ra parte (Figura 4)	Se verifica que el archivo config.txt contenga 11 parámetros del tipo adecuado.
1.2	Definir parámetros 2da parte (Figura 5)	Se verifica que el archivo config.txt contenga 11 parámetros del tipo adecuado.
1.3	Set Parameters (Figura 6)	Permite continuar con la asignación de parámetros.

Asume	Condición	Clases Correctas	Clases Erróneas
A	Numero de parámetros	{N=11} 1	{{N≠ 11} 2

	B	Tipo de parámetros	{celdas ∈ int} 3 {secciones ∈ int } 4 {Columnas espacio. Co – evo. ∈ int} 5 {Filas espacio. Co – evo. ∈ int} 6 {tamaño de la población ∈ int} 7 {columnas algoritmo ∈ int } 8 {filas algoritmo ∈ int } 9 {tasa de selección ∈ float} 10 {tasa de cruce ∈ float } 11 {tasa de mutación ∈ float } 12 {selección dirección PDB ∈ string } 13	{ celdas ∉ int} .14.1 {secciones ∉ int } 14.2 {Columnas espacio. Co – evo. ∉ int} 14.3 {Filas espacio. Co – evo. ∉ int} 14.4 {tamaño de la poblacion ∉ int} 14.5 {columnas algoritmo ∉ int } 14.6 {filas algoritmo ∉ int } 14.7 {tasa de selección ∉ float} 14.8 {tasa de cruce ∉ float } 14.9 {tasa de mutación ∉ float } 14.10 {selección dirección PDB ∉ string } 14.11
A, B	C	Rango de campos	{valor de tasas>0 y valor de tasas <1} 15 {celdas>1 y celdas <100}16 {Secciones = números que no sean primos y secciones < 26.}17 {Columnas espacio. Co – evo >0 y Columnas espacio. Co – evo <1}18 {Filas espacio. Co – evo > y Filas espacio. Co – evo <}19 {Filas espacio. Co – evo > y Filas espacio. Co – evo <}20 {tamaño de la población =	{valor de tasas<0 y/o valor de tasas >1} 28.1 {celdas<1 y/o celdas >100} 28.2 {Secciones = números que sean primos y secciones > 25.}28.3 {Columnas espacio. Co – evo >0 y Columnas espacio. Co – evo <1}28.4 {Filas espacio. Co – evo > y Filas espacio. Co – evo <}28.5 {Filas espacio. Co – evo > y Filas espacio. Co – evo <}28.6

			<p>números que no sean primos y tamaño de la población < 26 }21</p> <p>{columnas algoritmo > y columnas algoritmo < }22</p> <p>{filas algoritmo > y filas algoritmo <}23</p> <p>{tasa de selección > 0,01 y tasa de selección <1}24</p> <p>{tasa de cruce > 0,01 y columnas algoritmo < 1}25</p> <p>{ tasa de mutación >0,01 y tasa de mutación <1}26</p> <p>{ selección dirección PDB = Dirección archivos PDB}27</p>	<p>{tamaño de la población = números que sean primos y tamaño de la población > 25 }28.7</p> <p>{columnas algoritmo > y columnas algoritmo < }28.8</p> <p>{filas algoritmo > y filas algoritmo <}28.9</p> <p>{tasa de selección <0,01 y tasa de selección > 1}28.10</p> <p>{tasa de cruce <0,01 y tasa de cruce > 1}28.11</p> <p>{ tasa de mutación <0,01 y tasa de mutación > 1}28.12</p> <p>{selección dirección PDB ≠ Dirección archivos PDB }28.13</p>
--	--	--	---	---

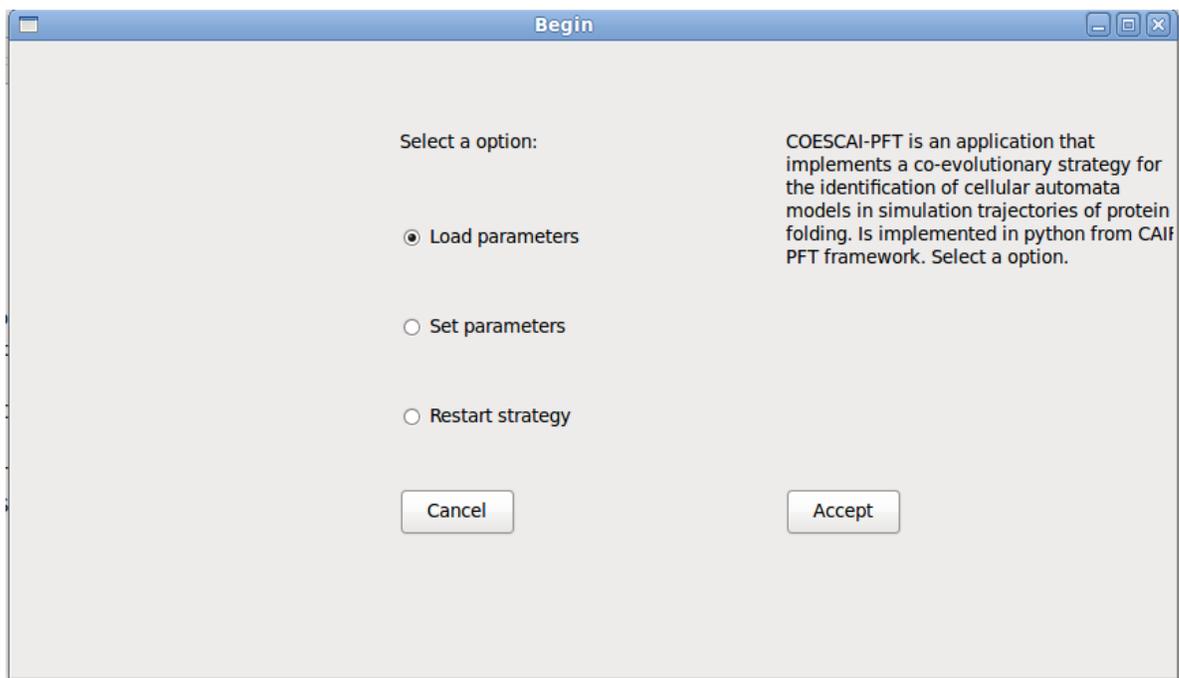


Figura 3. Interfaz Aplicación co- evolutiva, Begin

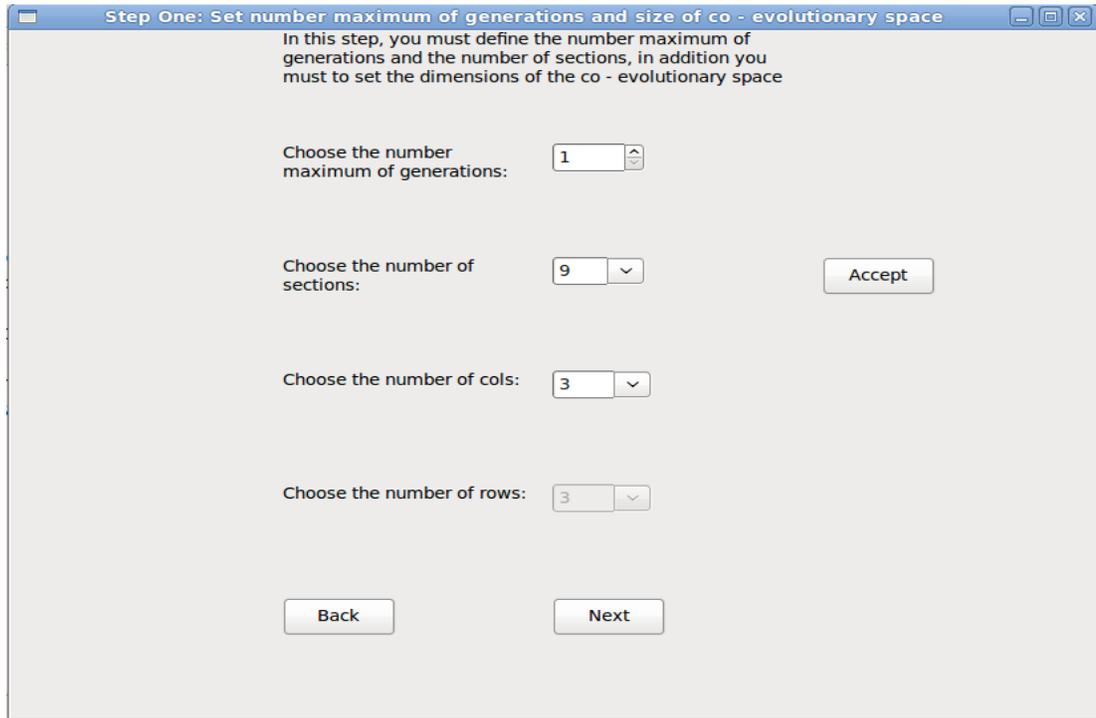


Figura 4. Interfaz Aplicación co- evolutiva, Definir parámetros 1ra parte.

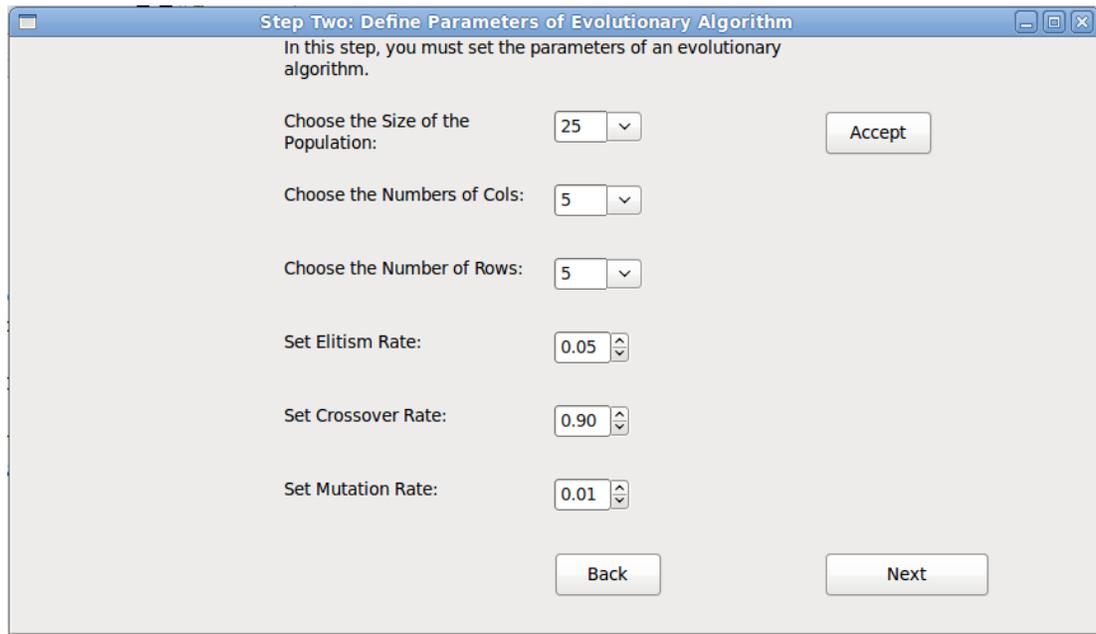


Figura 5. Interfaz Aplicación co- evolutiva, Definir parámetros 2da parte

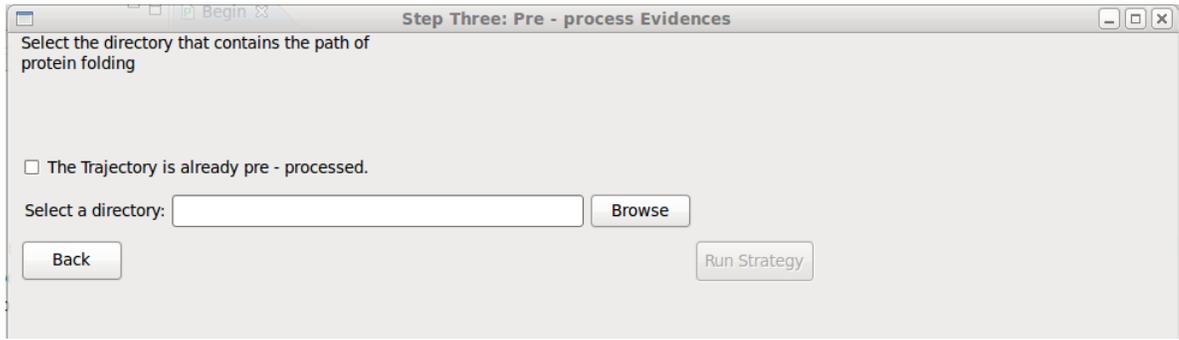


Figura 6. Interfaz Aplicación co- evolutiva, Set Parameters

2.2.2 Batería de Pruebas

	Entradas	Salidas esperadas	Clases cubiertas	Salida obtenida	Cumple
Clases correctas	(1,2,2,1,2,2,1,0.5,0.90,0.01, C:\Users\Adriana\Desktop\models-ContactMaps\)	Ejecución de estrategia	3,4,5,6,7, 8,9,10,11 ,12,13.	Ejecución de la estrategia	Si
	(1,2,2,1,2,2,1,0.7,0.80,0.05, C:\Users\Adriana\Desktop\models-ContactMaps\)	Ejecución de estrategia	3,4,5,6,7, 8,9,10,11 ,12,13.	Ejecución de la estrategia	Si
Clases erróneas	(1,a,2,1,2,2,1,1,1.90,1.01, C:\Users\Adriana\Desktop\models-ContactMaps\)	No debe permitir ir al siguiente paso, debería salir un mensaje de error.	3,4,5,6,7, 8,9,10,11 ,12,13, 28.10,28. 11,28.12.	No debe permitir ir al siguiente paso	NO
	(1,2,2,1,2,2,1,a,0.90,0.01, C:\Users\Adriana\Desktop\models-ContactMaps\)	No debe permitir ir al siguiente paso, debería salir un mensaje de error.	3,4,5,6,7, 8,9,10,11 ,12,13, 28.10.	Permite ir al siguiente paso y además , intenta ejecutar la estrategia	Si
	(1,2,2,1,2,2,1,0.5,a,0.01, C:\Users\Adriana\Desktop\models-ContactMaps\)	No debe permitir ir al siguiente	3,4,5,6,7, 8,9,10,11 ,12,13,28	Permite ir al siguiente paso y	Si

		paso, debería salir un mensaje de error.	.11.	además , intenta ejecutar la estrategia	
	(1,2,2,1,2,2,1,0.5,0,3,a, C:\Users\Adriana\Desktop\models-ContactMaps\)	No debe permitir ir al siguiente paso, debería salir un mensaje de error	3,4,5,6,7, 8,9,10,11 ,12,13,28 .12.	Permite ir al siguiente paso y además , intenta ejecutar la estrategia	Si
	(1,2,2,1,2,2,1,0.5,0,9,0.01, C:\Users\Adriana\Desktop\)	Debe salir un mensaje de error, que la dirección seleccionada , no contiene archivos PDB.	3,4,5,6,7, 8,9,10,11 ,12,13,28 .13.	No permite ejecutar la estrategia	Si
	(1,2,2,1,2,2,1,1,0,9,0.01, C:\Users\Adriana\Desktop\models-ContactMaps\)	Debe salir un mensaje de error, que la tasa de selección se pasa del rango establecido	3,4,5,6,7, 8,9,10,11 ,12,13, 28,10.	Permite ir al siguiente paso y además , intenta ejecutar la estrategia	Si
	(1,2,2,1,2,2,1,0.5,1.5,0.01, C:\Users\Adriana\Desktop\models-ContactMaps\)	Debe salir un mensaje de error, que la tasa de cruce se pasa del rango establecido	3,4,5,6,7, 8,9,10,11 ,12,13,28 ,11.	Permite ir al siguiente paso y además , intenta ejecutar la estrategia	Si
	(1,2,2,1,2,2,1,0.5,0,9,1,5, C:\Users\Adriana\Desktop\models-ContactMaps\)	Debe salir un mensaje de error, que la tasa de mutación se pasa del rango establecido	3,4,5,6,7, 8,9,10,11 ,12,13,28 ,12.	Permite ir al siguiente paso y además , intenta ejecutar la estrategia	Si

2.2.3 Historia de Revisiones

Fecha	Versión	Descripción	Autor
<2011-11-23>	<1.0>	<Pruebas de caja negra >	<Adriana Gómez>

Si desea ver más detalles sobre la implementación, ver en anexos manual técnico.