



**Universidad
Pontificia
Bolivariana**

SECCIONAL BUCARAMANGA

Minería de datos mediante la técnica de agrupamiento “cluster analysis” para un conjunto de datos de pacientes con indicios de enfermedades coronarias

José Joya Bulla

Christian Fernando Velandia

Análisis y diseño de algoritmos

Desarrollo de Software Orientado a WEB

Ingeniería de Software II

Telemática

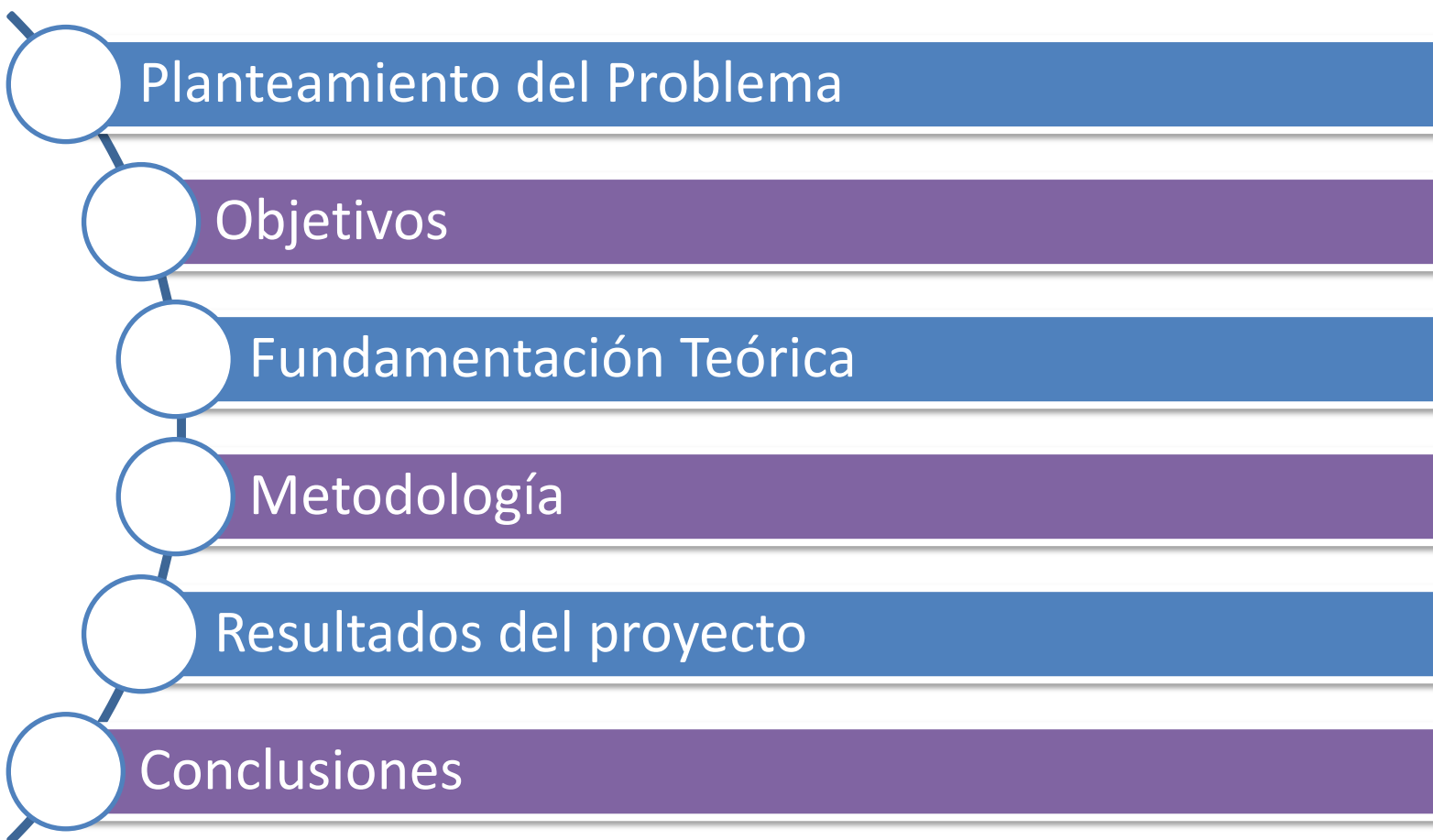
Sistemas Operativos



5a

Jornada de Socialización de Proyectos Integradores
Facultad de Ingeniería Informática

Agenda

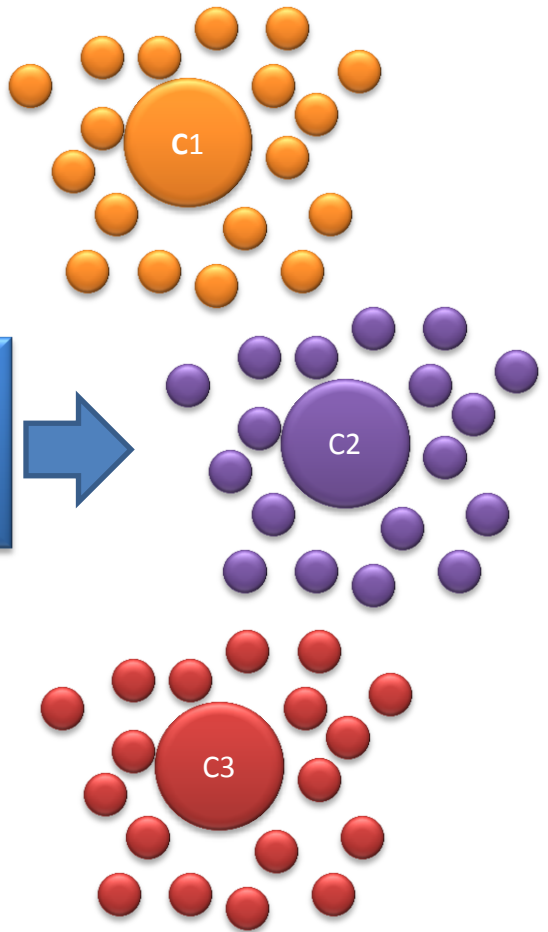


Planteamiento de la Situación Problema

```
@relation cleveland-14-heart-disease
@attribute 'age' real
@attribute 'sex' { female, male}
@attribute 'cp' { typ_angina, asympt, non_anginal, atyp_angina}
@attribute 'trestbps' real
@attribute 'chol' real
@attribute 'fbs' { t, f}
@attribute 'restecg' { left_vent_hyper, normal, st_t_wave_abnormality}
@attribute 'thalach' real
@attribute 'exang' { no, yes}
@attribute 'oldpeak' real
@attribute 'slope' { up, flat, down}
@attribute 'ca' real
@attribute 'thal' { fixed_defect, normal, reversable_defect}
@attribute 'num' { '<50', '>50_1', '>50_2', '>50_3', '>50_4'}
@data
63,male,typ_angina,145,233,t,left_vent_hyper,150,no,2.3,down,0,fixed
67,male,asympt,160,286,f,left_vent_hyper,108,yes,1.5,flat,3,normal,'>
67,male,asympt,120,229,f,left_vent_hyper,129,yes,2.6,flat,2,reversabl
37,male,non_anginal,130,250,f,normal,187,no,3.5,down,0,normal,'<50'
41,female,atyp_angina,130,204,f,left_vent_hyper,172,no,1.4,up,0,norma
56,male,atyp_angina,120,236,f,normal,178,no,0.8,up,0,normal,'<50'
62,female,asympt,140,268,f,left_vent_hyper,160,no,3.6,down,2,normal,'
57,female,asympt,120,354,f,normal,163,yes,0.6,up,0,normal,'<50'
63,male,asympt,130,254,f,left_vent_hyper,147,no,1.4,flat,1,reversable
53,male,asympt,140,203,t,left_vent_hyper,155,yes,3.1,down,0,reversabl
57,male,asympt,140,192,f,normal,148,no,0.4,flat,0,fixed_defect,'<50'
```



CLUSTERING



Objetivos específicos

Llevar a cabo la toma de requerimientos para establecer los alcances del problema en un documento de especificación de requerimientos.

Implementar un algoritmo clásico de análisis y agrupamiento de la información suministrada.

Diseñar y desarrollar un algoritmo propio y comparar el desempeño de los algoritmos.



5a

Jornada de Socialización de Proyectos Integradores
Facultad de Ingeniería Informática

Objetivos específicos



Aplicar el concepto de multiproceso para generar algoritmo aislado empleando varios procesos simultáneos.

Generar el diseño de la arquitectura de red para configurar y asignar direccionamiento a los dispositivos de la misma.



5a

Fundamentación Teórica

- Ingeniería de Requerimientos.
- Mejoramiento del proceso del Software.

- Algoritmos clásicos.
- Programación dinámica.
- GK-Modes



- HTML5, PHP, JAVASCRIPT
- Modelo Vista Controlador.

- Multiprocesos
- Diseño de Arquitectura de red.



METODOLOGÍA MODELO EN CASCADA

Análisis y definición de
requerimientos



Diseño del sistema y del
software.



Implementación y prueba de
unidades.

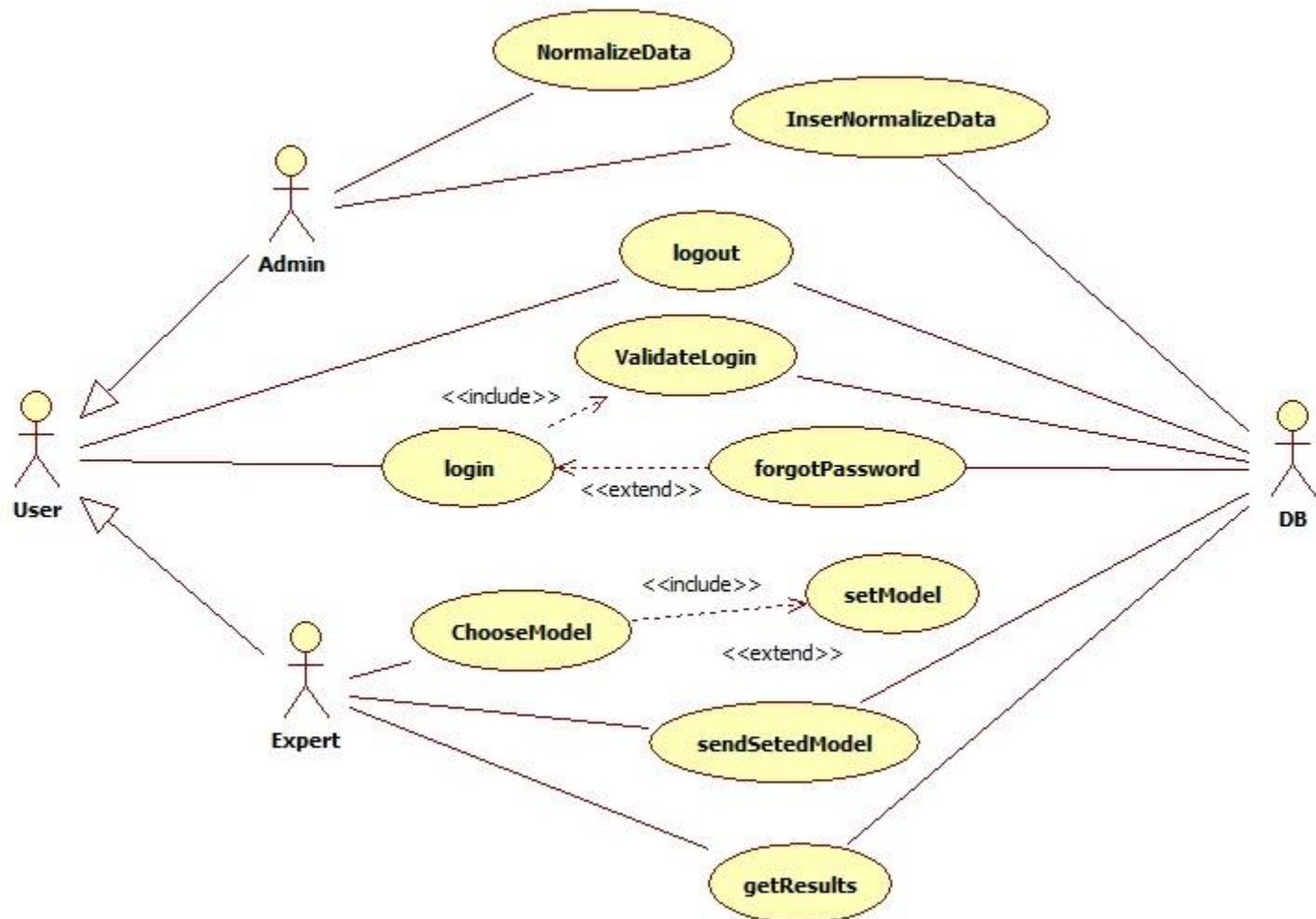


Integración y prueba del sistema



5a

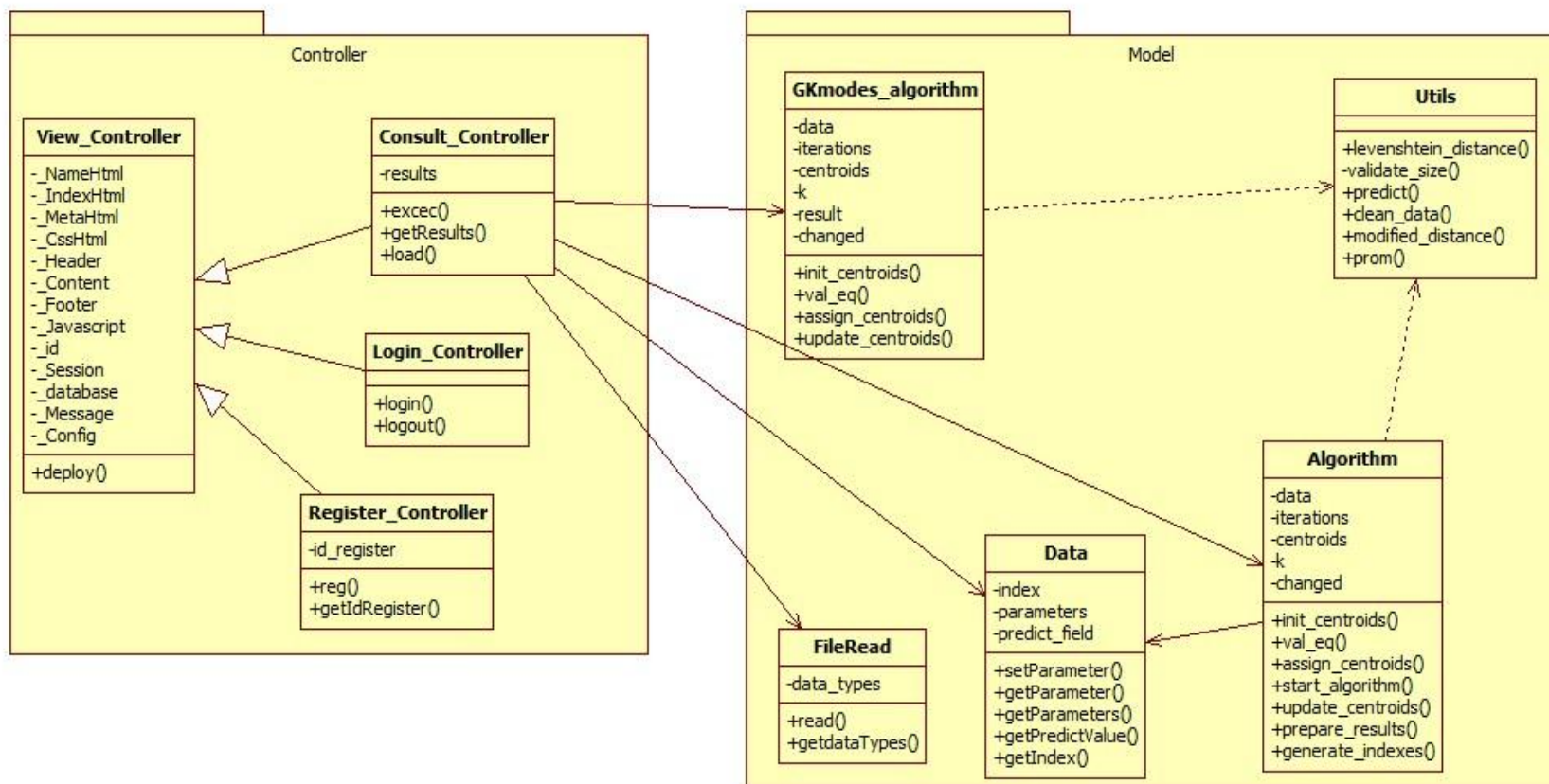
Resultados del Proyecto



5a

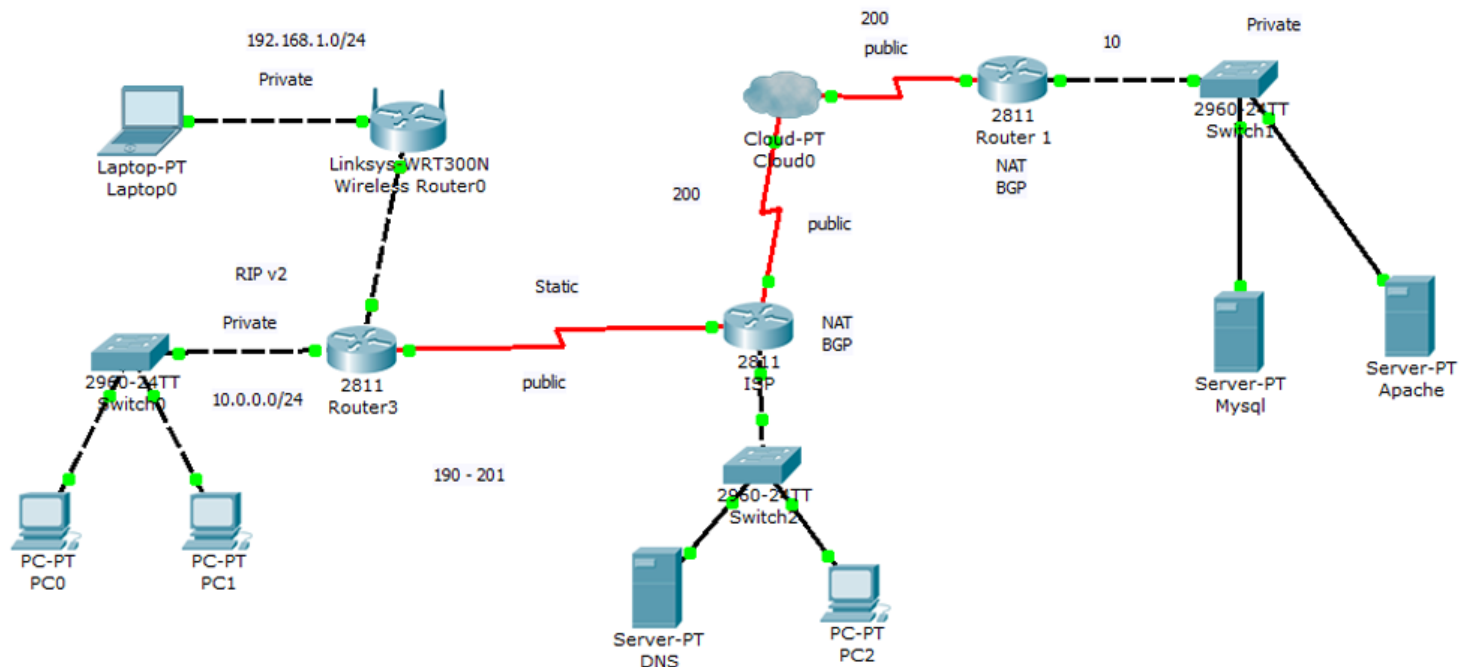
Resultados del Proyecto

Jornada de Socialización de Proyectos Integradores
Facultad de Ingeniería Informática

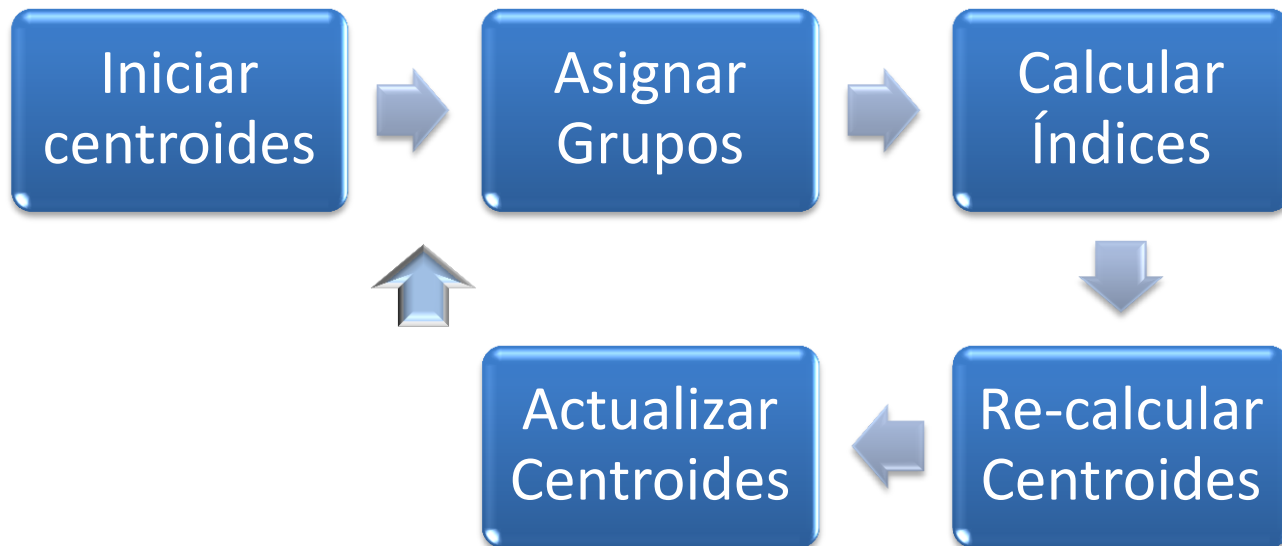


5a

Resultados del Proyecto



Resultados del Proyecto




5a

Resultados del Proyecto

DMUPB Login - Google Chrome


test.local/DMUPB/public/index.php?controller=login

Outlook Moodle UPB-B... SIIGA wowhead ubuntu-guia Bienestar Android Developers W3Schools MEGA TemplatesWEB



DMUPB
Data mining UPB

[INGRESAR](#) [HOME](#) [DOCUMENTACIÓN](#)



DMUPB
Data mining UPB

Universidad Pontificia Bolivariana
Seccional Bucaramanga

Usuario

Contraseña

[Ingresar](#)

Powered by Jose Joya Bulla, Christian Fernando Velandia
Proyecto Integrador I Semestre 2013

5a


Resultados del Proyecto

DMUPB Register - Google Chrome

test.local/DMUPB/public/index.php?controller=register

Outlook Moodle UPB-B... SIIGA wowhead ubuntu-guia Bienestar Android Developers W3Schools MEGA TemplatesWEB

Formulario de Registro de Usuarios



Por favor, llenar el formulario que se encuentra a continuación. Al registrar un nuevo usuario, éste tendrá el derecho a cargar archivos de datos en formato arff y realizar su respectivo analisis, permitiendo encontrar similitudes a través de cifras y graficas de dispersión.

Nombre

Apellido

Cuenta

Password

Re-type Password

Telefono

email

Tipo de Usuario
Seleccionar... ▼

Registrarse

5a

Resultados del Proyecto

DMUPB Workspace - Google Chrome

test.local/DMUPB/public/index.php?controller=consult&method=exec#

Outlook Moodle UPB-B... SIIGA wowhead ubuntu-guia Bienestar Android Developers W3Schools MEGA TemplatesWEB

Información del Centriode 1

AGE: 54 **SEX:** male **CP:** non_anginal **THALACH:** 152 **EXANG:** no
TRESTBPS: 125 **CHOL:** 273 **FBS:** f **OLDPEAK:** 0.5 **SLOPE:** down **CA:** 1
RESTECG: left_vent_hyper **THAL:** normal **NUM:** <50

Distancia Intra/Extra-cluster

[1] RESPECTO A [2] = 2.93 **INTRA CLUSTER =** 1.48
[1] RESPECTO A [3] = 2.07
[1] RESPECTO A [4] = 3.29
[1] RESPECTO A [5] = 2.21
[1] RESPECTO A [6] = 4.07

AGE	SEX	CP	TRESTBPS	CHOL	FBS	RESTECG	THALACH	EXANG	OLDPEAK	SLOPE	CA	THAL	NUM
63	male	typ_angina	145	233	t	left_vent_hyper	150	no	2.3	down	0	fixed_defect	<50
67	male	asympt	160	286	f	left_vent_hyper	108	yes	1.5	flat	3	normal	>50_1
41	female	atyp_angina	130	204	f	left_vent_hyper	172	no	1.4	up	0	normal	<50
62	female	asympt	140	268	f	left_vent_hyper	160	no	3.6	down	2	normal	>50_1
56	female	atyp_angina	140	294	f	left_vent_hyper	153	no	1.3	flat	0	normal	<50
64	male	typ_angina	110	211	f	left_vent_hyper	144	yes	1.8	flat	0	normal	<50
58	female	typ_angina	150	283	t	left_vent_hyper	162	no	1	up	0	normal	<50
58	male	atyp_angina	120	284	f	left_vent_hyper	160	no	1.8	flat	0	normal	>50_1
65	female	non_anginal	140	417	t	left_vent_hyper	157	no	0.8	up	1	normal	<50
53	male	non_anginal	130	197	t	left_vent_hyper	152	no	1.2	down	0	normal	<50
44	male	atyp_angina	130	219	f	left_vent_hyper	188	no	0	up	0	normal	<50
54	male	non_anginal	125	273	f	left_vent_hyper	152	no	0.5	down	1	normal	<50
51	male	typ_angina	125	213	f	left_vent_hyper	125	yes	1.4	up	1	normal	<50
46	female	non_anginal	142	177	f	left_vent_hyper	160	yes	1.4	down	0	normal	<50

5a

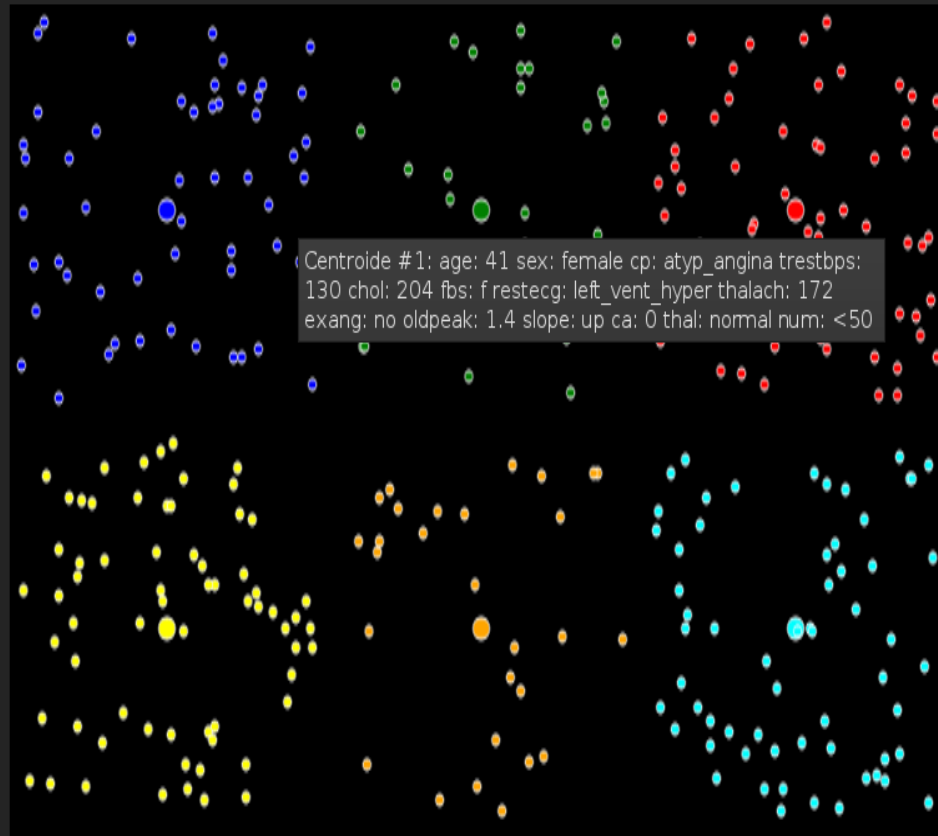
Jornada de Socialización de Proyectos Integradores
Facultad de Ingeniería Informática

Resultados del Proyecto

Algoritmo GK-MODES

Gráfica

Indice General Entre Clusters: 1.1726783110521



Conclusiones

- Los algoritmos que procesan datos mixtos consumen ligeramente mas recursos que aquellos que procesan números solamente.
- La complejidad temporal del “Algoritmo aislado” ($N * K * \log(N)$) es mayor al “GK-modes” ($K * \log(N)$) al tener que revisar cada parámetro como un valor por aparte.
- El uso de multiprocesos agiliza procedimientos que requieren de alto procesamiento de información.



Referencias Bibliográficas

- [1] R.C.T. Lee, S.S. Tseng, R.C. Chang, Y.T. Tsai. Introducción al Diseño y Análisis de Algoritmos. Un Enfoque Estratégico. McGraw-Hill 2007.
- [2] W3SCHOOLS.COM[en línea] <http://www.w3schools.com> [2012]
- [3] k. krishna and M. Murty (1999), 'Genetic K-Means algorithm', IEEE Transactions on Systems, Man, and Cybernetics Vol. 29, NO. 3, pp. 433-439.
- [4] Y. Lu, S. Lu, F. Fotouhi, Y. Deng, And S. Brown (2004), 'FGKA: A Fast Genetic K-Means Clustering Algorithm', ACM 1-58113-812-1.
- [5] International Journal Of Artificial Intelligence & Applications [en línea] airccse.org/journal/ijaia/papers/0410ijaia3.pdf [Apr. 2010]

