

## Un caso de estudio sobre la identificación de valores umbrales para medidas de código



**Carlos López, Esperanza Manso, Yania Crespo**

**GIRO** Grupo de Investigación en Reutilización y Orientación a Objeto

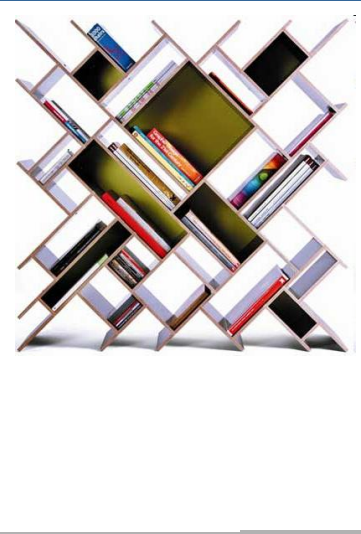


Universidad de Valladolid y Burgos

[clopezno@ubu.es](mailto:clopezno@ubu.es)

[{manso,yania}@infor.uva.es](mailto:{manso,yania}@infor.uva.es)

# Índice



- Contexto del problema
- Problema
- Solución propuesta
- Estudio empírico
- Conclusiones y líneas futuras



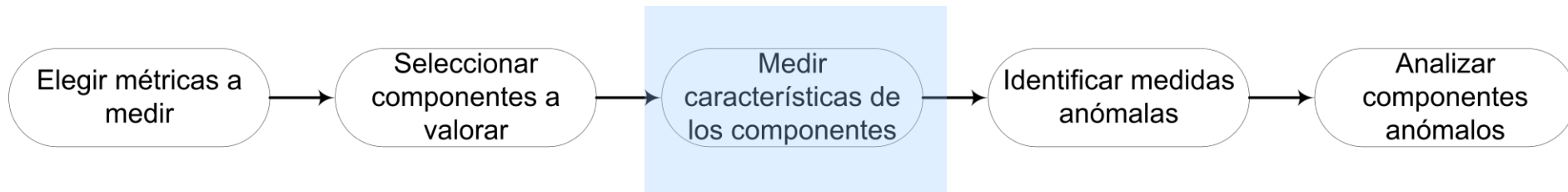


# Contexto del problema

- **Contexto del problema**
- Problema
- Solución propuesta
- Estudio empírico
- Conclusiones y líneas futuras

3/21

- **Medición del Software**
  - ***Actividad que forma parte de un proceso que consiste en asociar valores numéricos a atributos de productos, procesos y recursos***
- **Proceso de medición según Sommerville**





# Contexto del problema

## ■ Contexto del problema

- Problema
- Solución propuesta
- Estudio empírico
- Conclusiones y líneas futuras

4/21

## ¿Cómo identificar componentes anómalos?

### ■ Identificando medidas anómalas

#### ■ Obtener valores umbrales

- Estadísticos de agrupación
  - Media aritmética, Mediana, Cuartiles, etc.

#### ■ Valores umbrales absolutos vs. umbrales relativos a las aplicaciones

#### ■ Críticas sobre los umbrales :

- Están restringidos al contexto de medición
- No son aplicables a todas las entidades del contexto





# Contexto del problema

## ■ Contexto del problema

- Problema
- Solución propuesta
- Estudio empírico
- Conclusiones y líneas futuras

5/21

- Diseño de entidades de código
  - Categorizan las entidades por tipo de responsabilidad, según su **naturaleza**
  - Existen estereotipos UML estándar con este objetivo
    - *Utility, Entity, Interfaz, Control, Test...*
  - Los diseñadores y programadores utilizan esta clasificación en sus diseños





# Problema

- Contexto del problema
- **Problema**
- Solución propuesta
- Estudio empírico
- Conclusiones y líneas futuras

6/21

La información relativa a la naturaleza de la entidad no se utiliza en el proceso de medición para identificar medidas anómalas

## ■ Objetivos

- Mejorar el proceso de medición para incorporar la naturaleza de la entidad a medir
- Estudiar como influye la naturaleza de las entidades de código en los valores umbrales



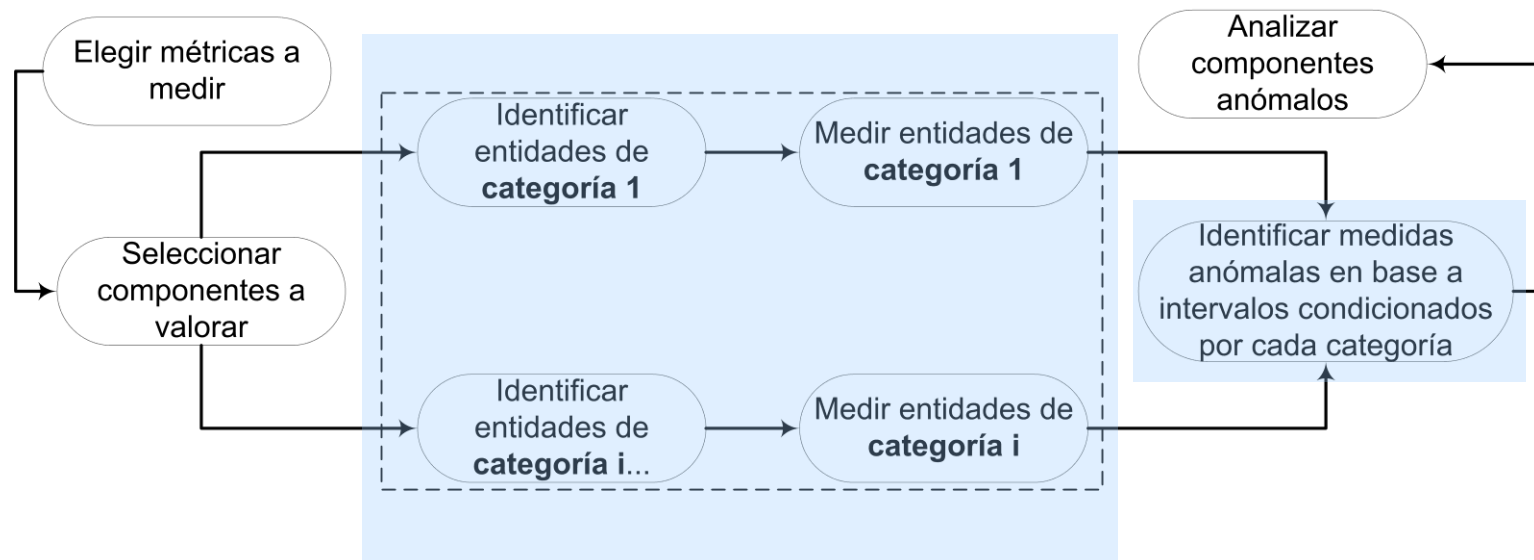


# Solución propuesta

## ■ Añadir una actividad de clasificación al proceso de medición

- Contexto del problema
- Problema
- **Solución propuesta**
- Estudio empírico
- Conclusiones y líneas futuras

7/21





# Solución propuesta

- Contexto del problema
- Problema
- **Solución propuesta**
- Estudio empírico
- Conclusiones y líneas futuras

8/21

- Realizar un caso de estudio empírico para validar la nueva propuesta

- Hipótesis

**Las métricas de entidades de código se comportan igual, independientemente de la naturaleza de la entidad**

- Las ***variables dependientes*** son las medidas de entidades de código de ( $M_i$ )
- La ***variable independiente*** es la clasificación de dichas entidades, basada en la siguiente escala nominal:

$e_1$  exception,  $e_2$  interface,  $e_3$  entity,  $e_4$  control,  $e_5$  test,  $e_6$  utility





# Estudio empírico Planificación

- Contexto del problema
- Problema
- Solución propuesta
- **Estudio empírico**
- Conclusiones y líneas futuras

9/21

## ■ Objetos y sujetos

### ■ Sujetos

- Dos estudiantes de 5º de Ingeniería Informática
  - Miden y clasifican entidades de código
- Un profesor tutor de su TFC
  - Clasifica entidades

### ■ Objetos

- Códigos fuente obtenidos del repositorio *SourceForge*
- Se utilizan criterios de selección de objetos
- 10 plugins de Eclipse – según criterios
- Todos los sujetos participan en la clasificación de todos los objetos para consensuar una **Tabla de Convención de Nombres** hasta conseguir 50% clasificación



# Estudio empírico Planificación

- Contexto del problema
- Problema
- Solución propuesta
- **Estudio empírico**
- Conclusiones y líneas futuras

10/21

## ■ Instrumentación

- *RefactorIt*
  - Herramienta de cálculo de medidas
- Clasificación de medidas por característica
  - tamaño (**TAM**)
  - documentación (**DOC**)
  - acoplamiento (**ACO**)
  - herencia (**HER**)
  - complejidad estructural (**COM**)
  - Abstracción (**ABS**)
  - Cohesión (**COH**)
  - Principios de diseño (**PDA**)
- Clasificación por ámbito
  - **P** paquete
  - **C** Clase
  - **M** Método

Descripción	Identificador	Min Valor	Max Valor	Ámbito	Característica
Comment Lines of Code	CLOC			T	TAM
Cyclomatic Complexity	V(G)	1	10	M	COM
Density of Comments	DC	0.2	0.4	T	DOC
Executable Statements	EXEC	0	20	T	TAM
Non-Comment Lines of Code	NLOC			T	TAM
Number of Parameters	NP	0	4	M	TAM
Total Lines of Code	LOC	5	1000	T	TAM
Abstractness	A	0.0	0.5	P	ABS
Afferent Coupling	Ca	0	500	P	ACO
Depth in Tree	DIT	0	5	C	HER
Efferent Coupling	Ce	0	20	P	ACO
Instability	I	0.7	1.0	P	ACO
Number of Abstract Types	NOTa	0	20	P	ABS
Number of Children	NOC	0	10	C	HER
Number of Concrete Types	NOTc	0	80	P	ABS
Number of Exported Types	NOTe	3	50	P	ACO
Number of Fields	NOF	0	1	C	TAM
Number of Types	NOT	0	80	P	TAM
Response for Class	RFC	0	50	C	COM
Weighted Methods per Class	WMC	1	50	C	COM
Number of Attributes	NOA	0	5	C	TAM
Cyclic Dependencies	CYC	0	1	P	PDA
Dependency Inversion Principle	DIP	0.3	1.0	C	PDA
Direct Cyclic Dependencies	DCYC	0	1	P	PDA
Distance from the Main Sequence	D	0.0	0.1	P	PDA
Encapsulation Principle	EP	0	0.6	P	PDA
Lack of Cohesion of Methods	LCOM	0.0	0.2	C	COH
Limited Size Principle	LSP	0	10	P	PDA
Modularization Quality	MQ	0	1000	P	PDA
Number of Tramps	NT	0	1	M	O



# Estudio empírico Operación

- Contexto del problema
- Problema
- Solución propuesta
- **Estudio empírico**
- Conclusiones y líneas futuras

11/21

## ■ Operación

- Algoritmo de clasificación que se basa en criterios de convención de nombres de las entidades
  - Precondición: Se dispone de una **Tabla de Convención de Nombres**
- 1. Obtener el *nombre simple* y *nombre cualificado* de la entidad de código
- 2. Buscar si el *nombre* esta contenido en la **Tabla de Convención de Nombres**
  - Asocia cadenas de caracteres a categorías de entidades
  - Ejemplo: "Action" => e4 Entidad de control
- 3. En el caso de **conflicto**, porque el nombre contenga cadenas que correspondan a más de una categoría
  - Prevalece el criterio del nombre simple de la entidad



# Estudio empírico Operación

- Contexto del problema
- Problema
- Solución propuesta
- **Estudio empírico**
- Conclusiones y líneas futuras

12/21

## ■ Operación

### ■ Tabla de Convención de Nombres

Categorías	e <sub>1</sub>	e <sub>2</sub>	e <sub>3</sub>	e <sub>4</sub>	e <sub>5</sub>	e <sub>6</sub>
Convención de nombres	exception	interface gui forms ui report swing visual view awt	core model entity	control facade manager handler action callback maker provider	test debug dummy	util properties log preference template options

## ■ Entradas y salidas esperadas del algoritmo

<code>com.mountains.eclemma.core.test</code>	<b>e5</b>	(conflicto)
<code>ui.actions.MergeSessionAction</code>	<b>e4</b>	(conflicto)
<code>core.analysis.JavaElementCoverage</code>	<b>e3</b>	
<code>core.SessionManager</code>	<b>e4</b>	(conflicto)
<code>ui.coverageview.CoverageView</code>	<b>e2</b>	



# Estudio empírico Operación

- Contexto del problema
- Problema
- Solución propuesta
- **Estudio empírico**
- Conclusiones y líneas futuras

13/21

## ■ Operación

### ■ Descripción de la clasificación de objetos

	Número de entidades							Información Sourceforge			Tamaño
Plugins Eclipse	e <sub>1</sub>	e <sub>2</sub>	e <sub>3</sub>	e <sub>4</sub>	e <sub>5</sub>	e <sub>6</sub>	Sin clasificar	%Actividad	Nº Descargas	Estado	LOC
esFtp	1	90	38	68	8	86	52	94,19	45878	4	5156
AVR	12	376	125	306	362	789	310	98,68	1464780	5,6	55004
Jedit	25	1671	1657	1548	12	1719	681	99,62	5371954	5,6	156656
EclEmma	1	257	288	78	186	35	150	99,93	1488061	5	13294
AzSMRC	45	511	655	272	9	759	758	99,00	59327	4,5	67760
EclipseME	39	575	951	535	281	1215	446	97,63	731177	5	80012
ELBE	26	1788	1561	1216	138	380	288	85,95	30172	7	76188
OpenReports	17	0	568	1074	2	159	383	96,80	235708	4,5	27338
EclipseCorba	7	145	148	232	143	205	334	94,86	23062	3	25051
LabelDecorator	7	0	116	34	52	177	74	85,00	7004	5	5256
<b>LOC Totales</b>	<b>909</b>	<b>111818</b>	<b>85911</b>	<b>83876</b>	<b>14238</b>	<b>14797</b>	<b>200166</b>				



# Estudio empírico Análisis

- Contexto del problema
- Problema
- Solución propuesta
- **Estudio empírico**
- Conclusiones y líneas futuras

14/21

## ■ Análisis de resultados

- La clasificación no es completa
  - 39,2 % de entidades sin clasificar
  - El algoritmo no es exhaustivo
  - Hay entidades que no se pueden clasificar

## ■ Hipótesis

**$H_0$ :** La métrica  **$M_i$**  se comporta igual en las entidades de código, independientemente de su ***naturaleza***

Donde

**$M_i$**  métricas de la tabla

***naturaleza***  $e_1$  *exception*,  $e_2$  *interface*,  $e_3$  *entity*,  $e_4$  *control*,  $e_5$  *test*,  $e_6$  *utility*



# Estudio empírico Análisis

- Contexto del problema
- Problema
- Solución propuesta
- **Estudio empírico**
- Conclusiones y líneas futuras

15/21

## ■ Análisis de resultados

### ■ Test ANOVA (**AN**alysis **Of** **VA**riance)

- Niveles de tratamiento: estereotipos UML

$$H_0 : \mu_{e1}^{M_i} = \mu_{e2}^{M_i} = \mu_{e3}^{M_i} = \mu_{e4}^{M_i} = \mu_{e5}^{M_i} = \mu_{e6}^{M_i}$$

- Se aplica ANOVA no paramétrico de Kruskal-Wallis (KW)
- Valores umbrales por categoría
  - Si se rechaza  $H_0$  entonces
    - Calcular umbrales (P25,P75)





# Estudio empírico Análisis

- Contexto del problema
- Problema
- Solución propuesta
- **Estudio empírico**
- Conclusiones y líneas futuras

16/21

- P-valores ANOVA (KW)
  - Significativos alfa **0.05**
  - Global
    - 34 de 39 **(87%)**
  - Ámbito
    - P 16 de 19 **(84%)**
    - C 9 de 11 **(82%)**
    - M 7 de 7 **(100%)**
  - Características
    - COM 3 de 3 **(100%)**
    - TAM 12 de 12 **(100%)**
    - DOC 3 de 3 **(100%)**
    - ABS 3 de 3 **(100%)**
    - **ACOP** 3 de 4 **(75%)**
    - **HER** 1 de 2 **(50%)**
    - **PDA** 3 de 7 **(43%)**

Identificador	Característica	Resultados test K-W		
		Paquete	Clase	Método
NOT	TAM	0.02238	---	---
NP	TAM	---	---	< 2.2e-16
NOA	TAM	---	< 2.2e-16	---
NLOC	TAM	0.01340	< 2.2e-16	< 2.2e-16
EXEC	TAM	0.005242	< 2.2e-16	< 2.2e-16
CLOC	TAM	0.02767	< 2.2e-16	0.0129
MQ	PDA	0.1427	---	---
LSP	PDA	0.0775	---	---
EP	PDA	0.006352	---	---
D	PDA	0.0004983	---	---
DCYC	PDA	0.03573	---	---
DIP	PDA	---	0.2769	---
CYC	PDA	0.0002174	---	---
NT	O	---	---	< 2.2e-16
NOC	HER	---	0.2813	---
DIT	HER	---	< 2.2e-16	---
DC	DOC	0.01089	< 2.2e-16	0.0367
WMC	COM	---	< 2.2e-16	---
RFC	COM	---	2.68E-09	---
V(G)	COM	---	---	< 2.2e-16
LCOM	COH	---	< 2.2e-16	---
NOTe	ACO	0.1909	---	---
I	ACO	0.0001181	---	---
Ce	ACO	0.007674	---	---
Ca	ACO	0.003002	---	---
NOTc	ABS	0.001988	---	---
NOTa	ABS	0.03124	---	---
A	ABS	0.0005088	---	---





# Estudio empírico Análisis

- Contexto del problema
- Problema
- Solución propuesta
- **Estudio empírico**
- Conclusiones y líneas futuras

17/21

- Análisis de resultados
  - Valores umbrales por categoría
    - (P25,P75)
    - Interpretación umbrales WMC
      - Más preciso que recomendado por RefactorIt [1-50]
      - e1 y e5 intervalos muy precisos

		DC	LOC	NCLOC	EXEC	CLOC	WMC	DIT	RFC	LCOM	NOA
Exception e <sub>1</sub>	Q1	0.00	10.25	6.25	0	0	1.75	3	1.75	0	0
	Q3	0.14	29.25	15.00	1.75	3	4.00	4	3.25	0.00	2
Interfaz e <sub>2</sub>	Q1	0.00	35.00	24.00	2	0	4.00	1	2.00	0	1
	Q3	0.20	179.25	133.25	14.00	25	16.00	2	19.00	0.95	7
Entity e <sub>3</sub>	Q1	0.00	40.00	24.75	2	0.75	5.00	1	2.00	0	1
	Q3	0.27	169.50	119.25	17.00	32	25.00	2	21.00	0.96	6
Control e <sub>4</sub>	Q1	0.00	18.00	14.00	1	0	3.00	1	2.00	0	0
	Q3	0.21	115.25	79.00	12.00	13	16.25	2	17.00	0.75	3
Test e <sub>5</sub>	Q1	0.00	19.00	12.25	0	0	2.00	1	2.00	0	0
	Q3	0.28	105.00	71.25	8.00	19	8.25	2	9.00	0.81	2
Utility e <sub>6</sub>	Q1	0.02	33.00	22.00	2	1	4.00	1	3.00	0	1
	Q3	0.31	202.75	130.75	23.00	41	22.00	2	23.00	0.92	6



# Estudio empírico Validez

- Contexto del problema
- Problema
- Solución propuesta
- **Estudio empírico**
- Conclusiones y líneas futuras

18/21

- Existe una réplica del experimento con resultados similares RPME2009
  - Métricas -> Se amplía el conjunto de métricas
- Amenazas
  - Subjetividad de la clasificación
    - No es ortogonal
  - Eliminación de entidades por no poderse clasificar
  - No supervisión de los sujetos





# Conclusiones y líneas futuras

- Contexto del problema
- Problema
- Solución propuesta
- Estudio empírico
- **Conclusiones y líneas futuras**

19/21

Los resultados observados son favorables a considerar la naturaleza de las entidades cuando se trabaja con valores umbrales de métricas para detectar medidas anómalas

- Requerimientos para la transmisión de resultados a una Empresa
  - Clasificación de entidades de código
  - Histórico de medidas de entidades
  - Cálculo de valores umbrales para cada categoría





# Conclusiones y líneas futuras

- Contexto del problema
- Problema
- Solución propuesta
- Estudio empírico
- **Conclusiones y líneas futuras**

20/21

- **Más réplicas**
  - Nuevos objetos ...
  - Comparar valores umbrales
- **Validar el proceso de medición**
  - ¿Detecta mejor los componentes anómalas?
- **Validar el algoritmo de clasificación**
  - ¿ La técnica de clasificación basada en convención de nombre de las entidades podría sustituir a la clasificación de un experto?



## Gracias por su atención



**Carlos López, Esperanza Manso, Yania Crespo**

**GIRO** Grupo de Investigación en Reutilización y Orientación a Objeto



Universidad de Valladolid y Burgos

[clopezno@ubu.es](mailto:clopezno@ubu.es)

[{manso,yania}@infor.uva.es](mailto:{manso,yania}@infor.uva.es)