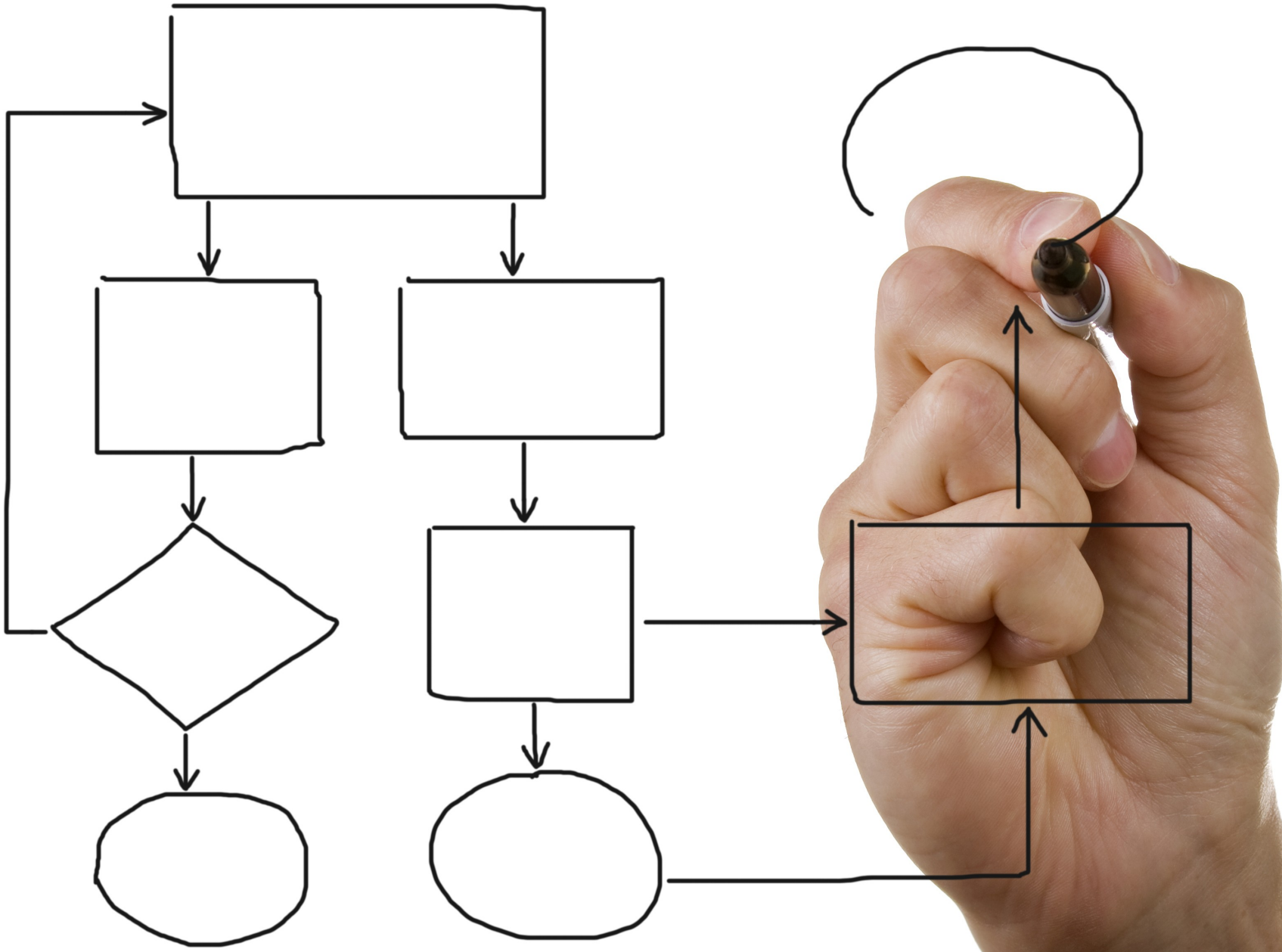
Multi-objective Module Clustering for Kate

Matheus Paixao, Mark Harman and Yuanyuan Zhang

{matheus.paixao.14, mark.harman, yuanyuan.zhang}@ucl.ac.uk

Centre for Research on Evolution, Search and Testing – CREST

Department of Computer Science – UCL

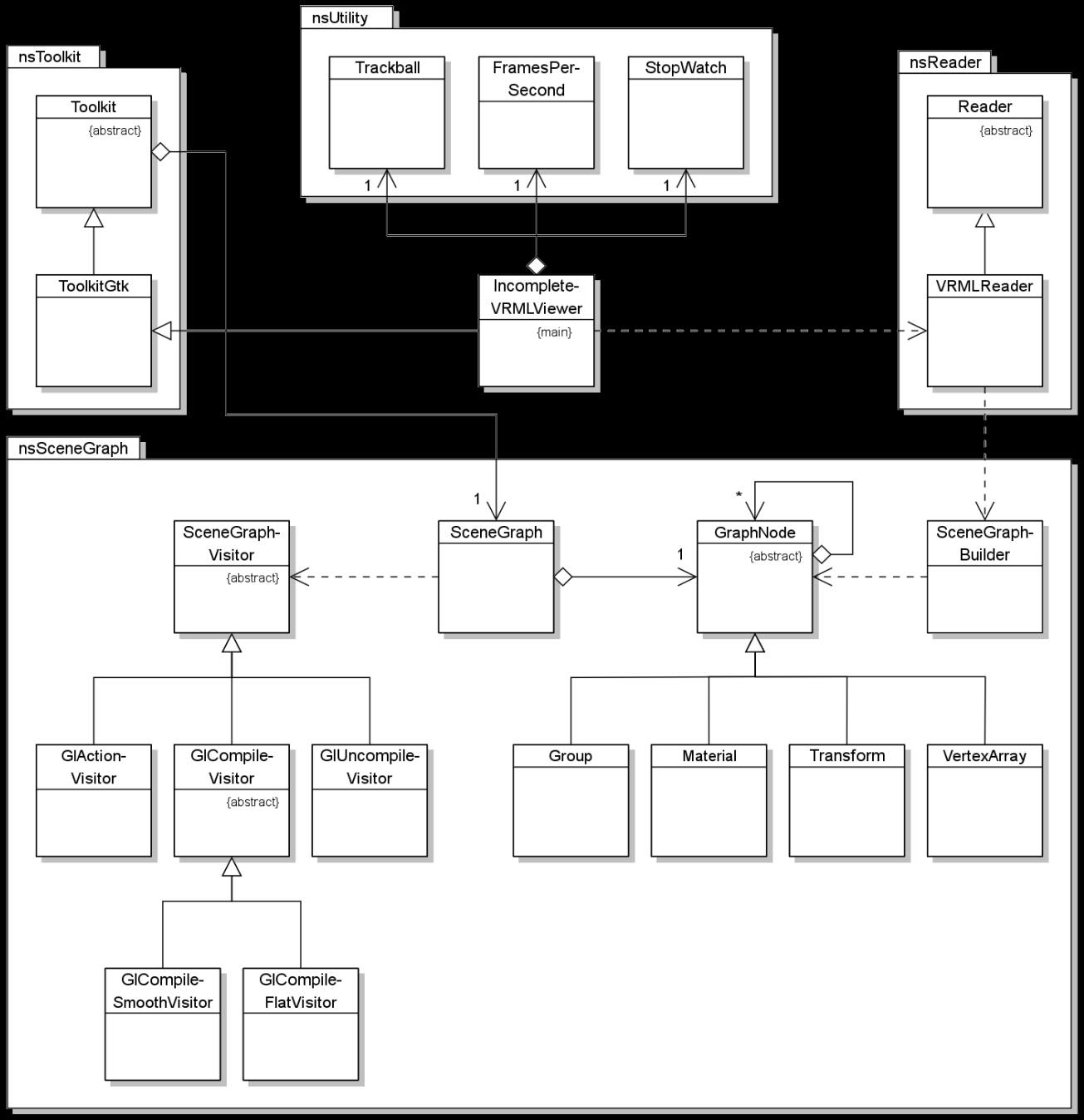


Project beggining :-)

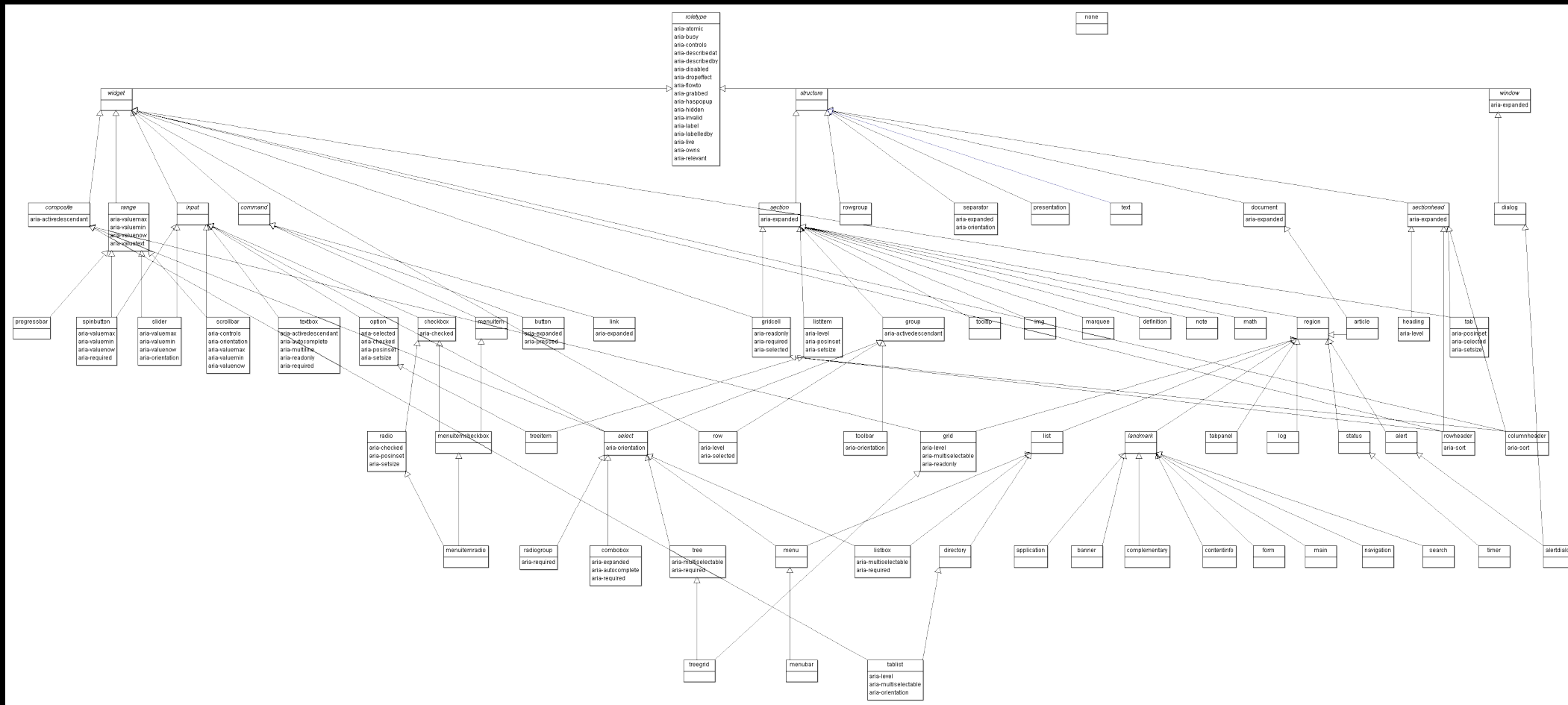
Develop

Evolve

Maintain



Project evolves :-(



Search Based Design Optimization

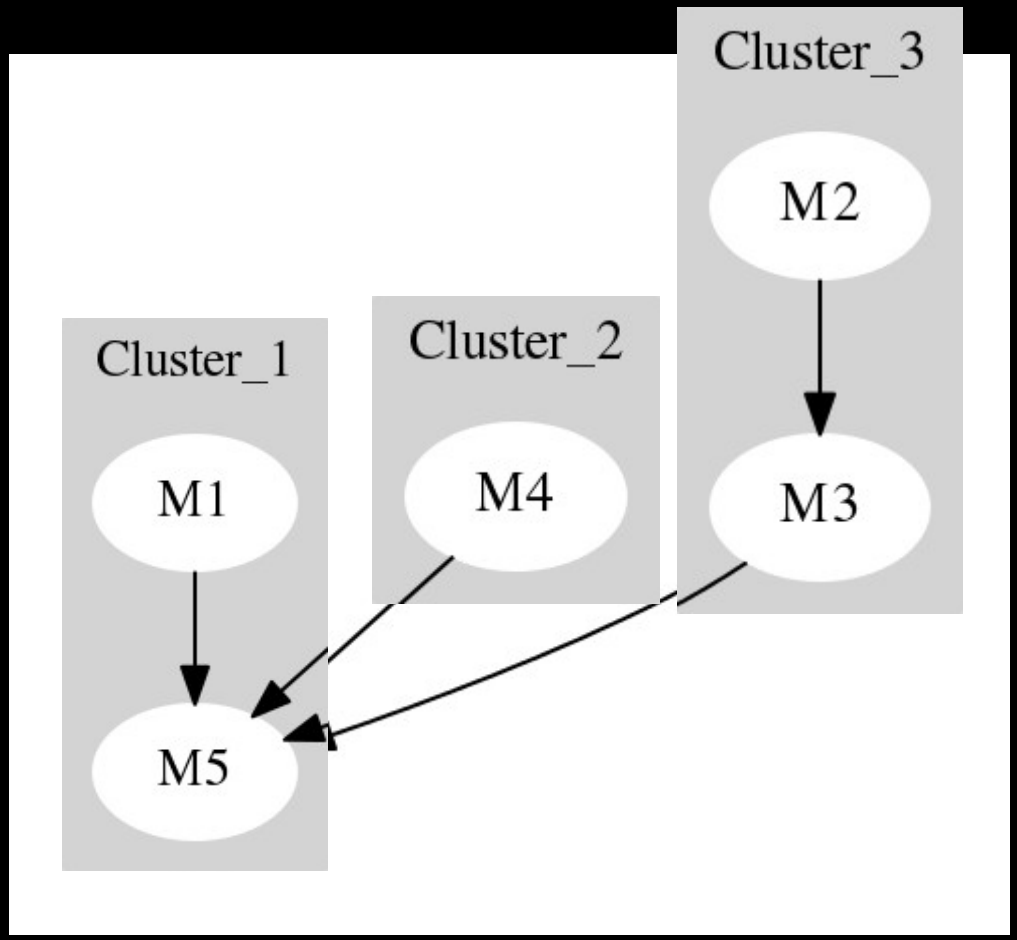
Design generation

Software Module Clustering

Improvement of existing design

Software Module Clustering

The system is decomposed in a set of modules

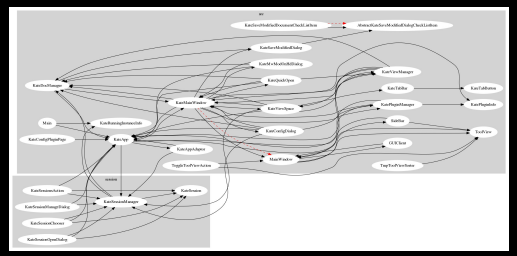


Multi-objecive Module Clustering for Kate

C/C++ text editor for KDE platforms

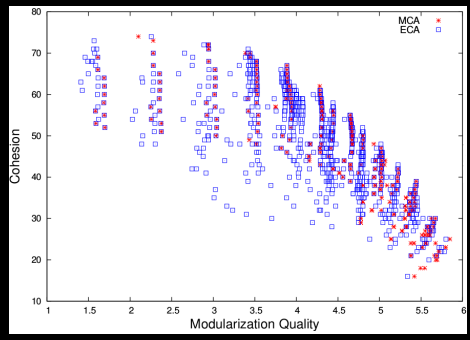
kate's source code

doxygen



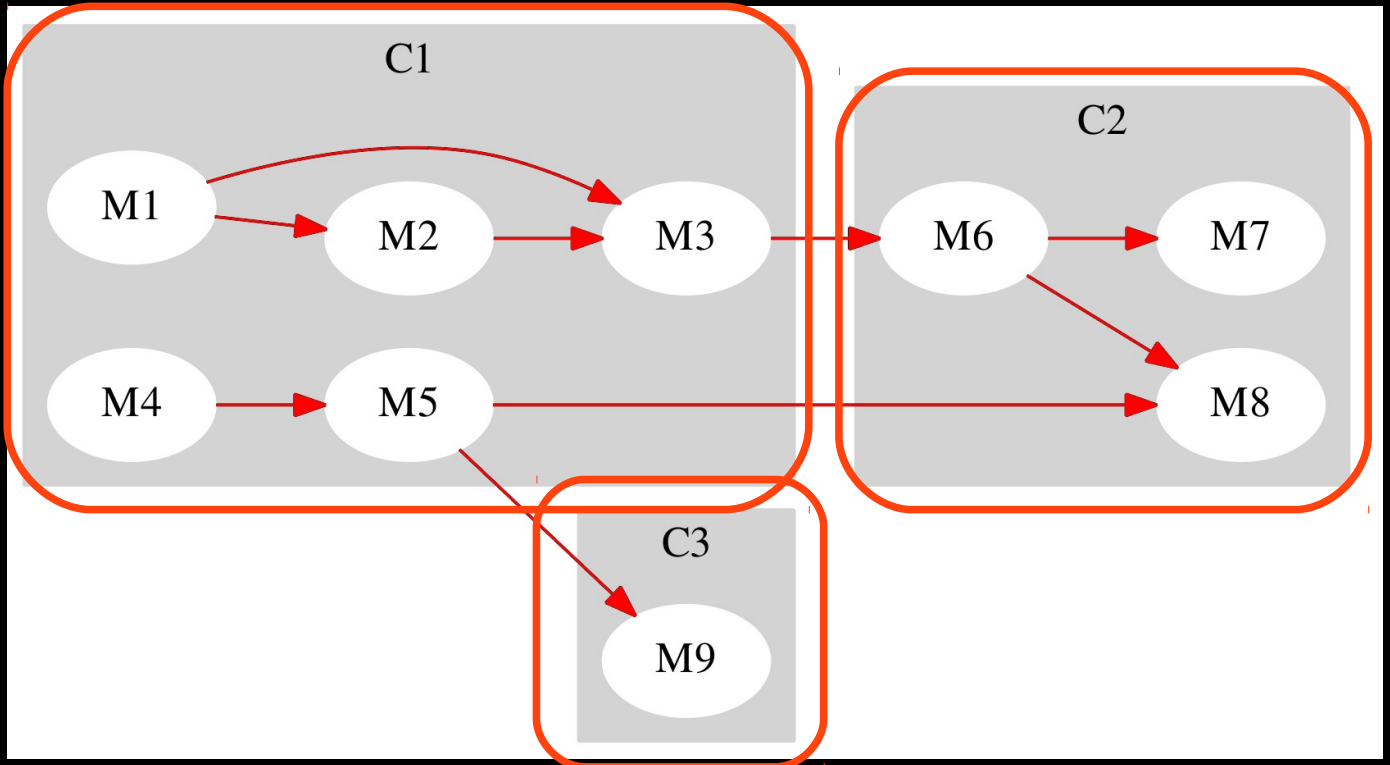
kate's original design

two-archive GA



pareto front of designs

Fitness Functions



cohesion: 6

coupling: 3

MQ: 1.38

number of clusters: 3

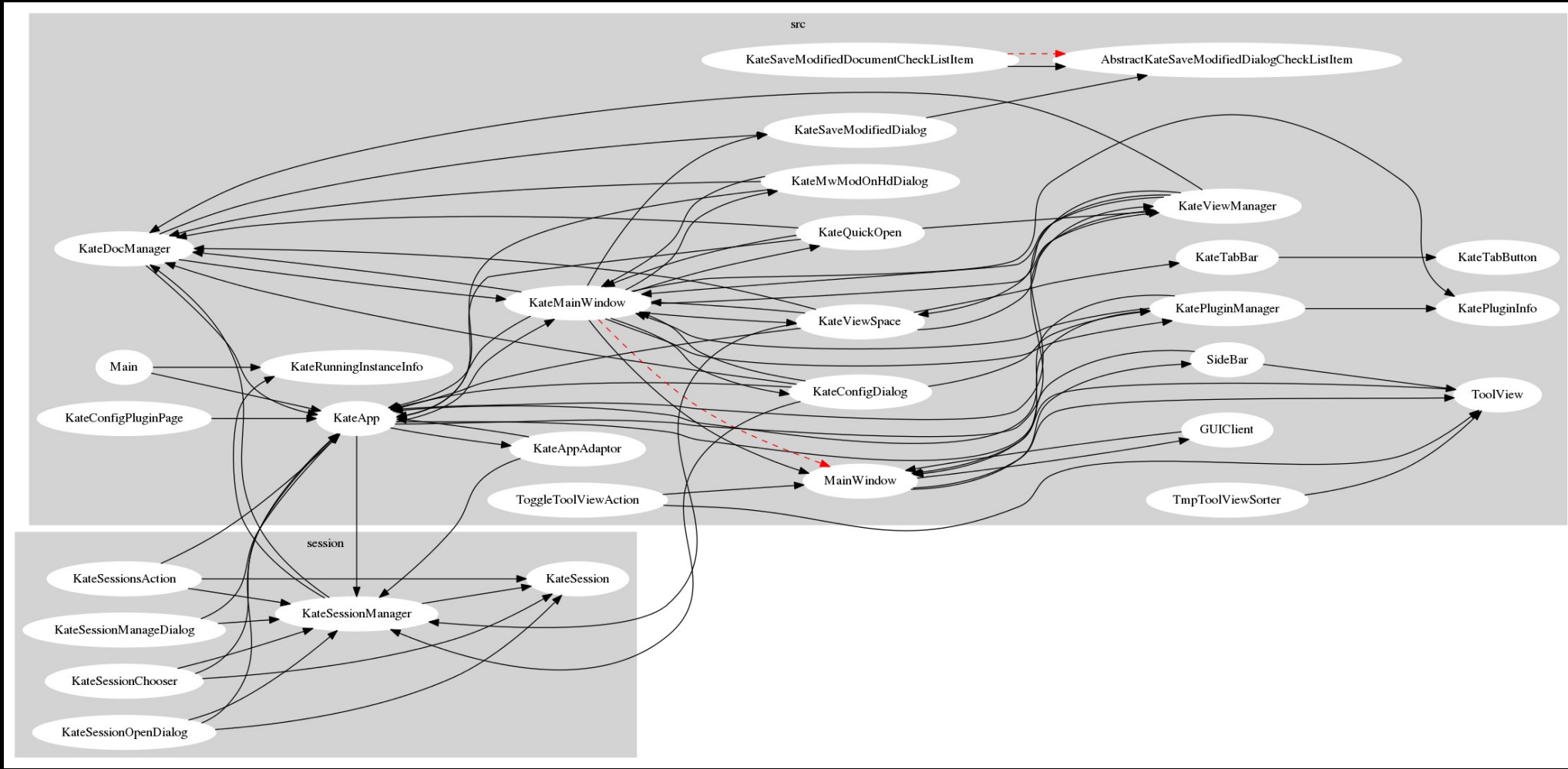
isolated clusters: 1

cluster size diff: 4

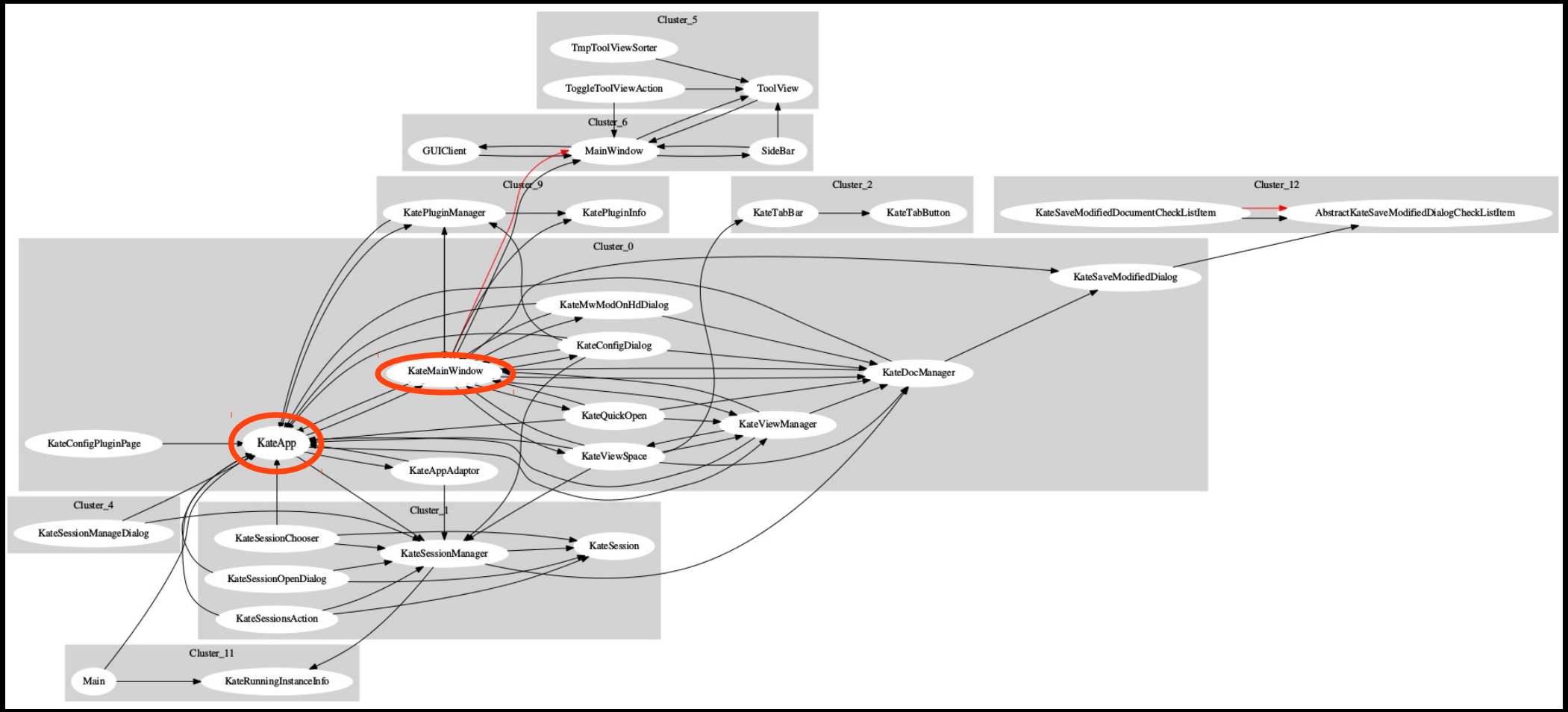
$$MQ = \sum_{k=1}^{N} MF(C_k)$$

$$\text{where, } MF(C_k) = \begin{cases} 0, & \text{if } i_k = 0 \\ \dfrac{i_k}{i_k + \frac{j_k}{2}}, & \text{if } i_k > 0 \end{cases}$$

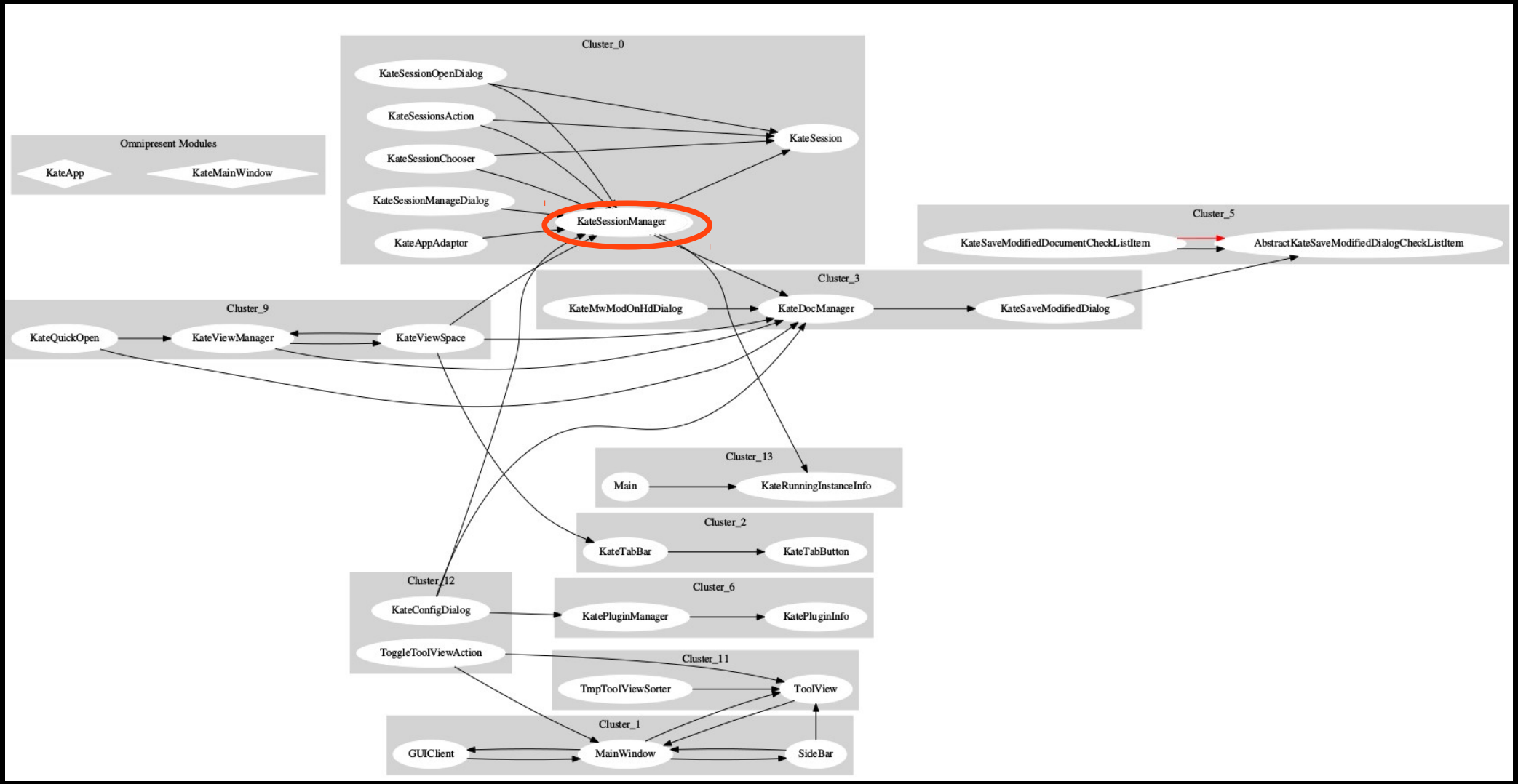
kate's original design



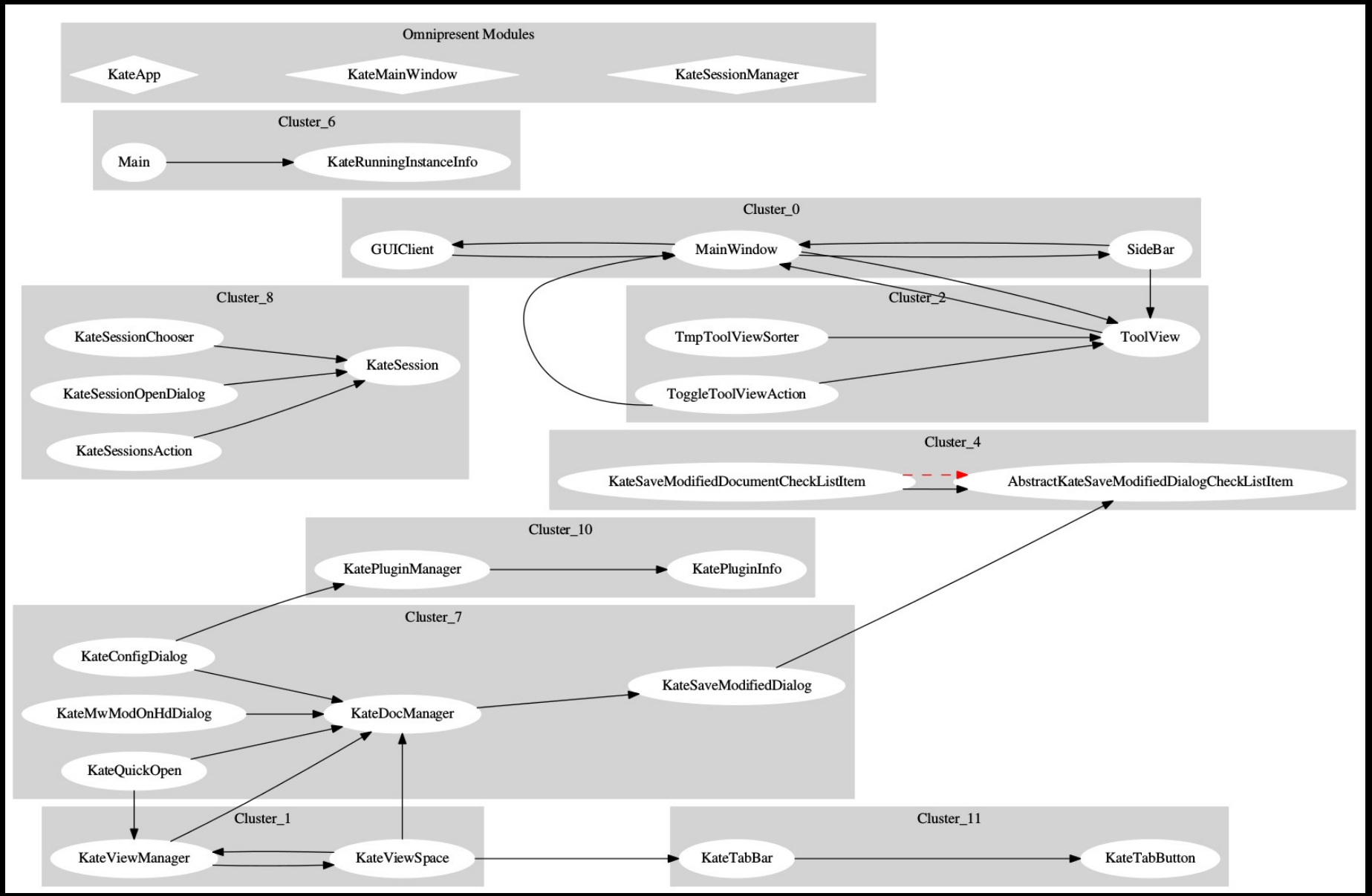
Omnipresent Modules



Results when omnipresent modules are considered



Results when omnipresent modules are considered



Limitations and Future Work

Handling omnipresent modules

Solutions visualization

Search process efficiency

For the Interested Reader

Technical report:

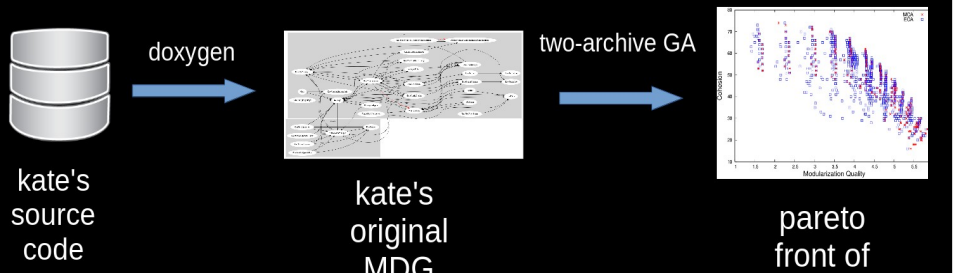
http://www.cs.ucl.ac.uk/research/research_notes/

Kate modularization datasets available

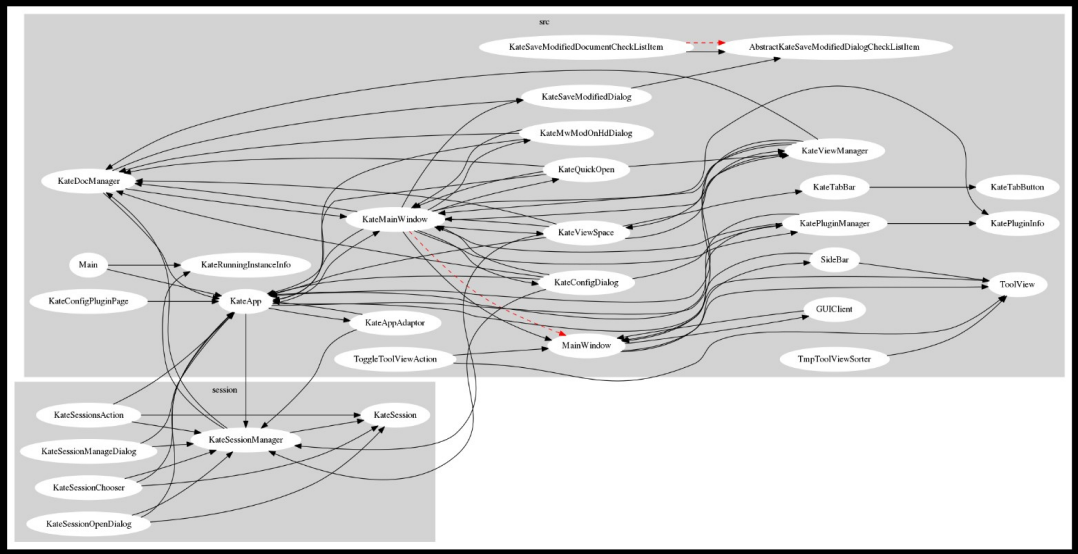
http://www0.cs.ucl.ac.uk/staff/m.paixao/kateMod/

Multi-objecive Module Clustering for Kate

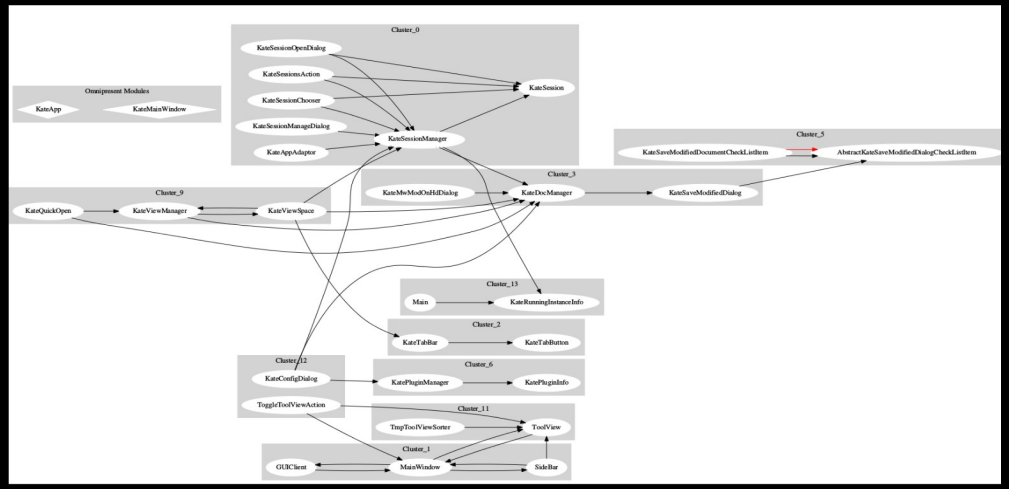
- C/C++ text editor for KDE platforms



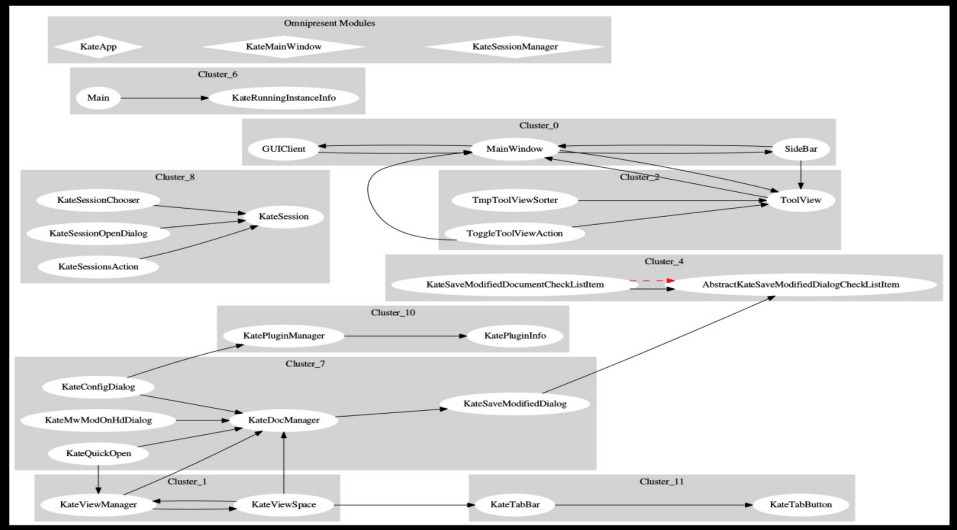
kate's original design



Results when omnipresent modules are considered



Results when omnipresent modules are considered



Fitness Functions Used

Maximize Cluster Approach (MCA)

- cohesion (max)
- coupling (min)
- number of clusters (max)
- isolated clusters (min)
- MQ (max)

Equal-size Cluster Approach (ECA)

- cohesion (max)
- coupling (min)
- number of clusters (max)
- clusters size difference (min)
- MQ (max)

GA Parameters

population size: 10M

generations: 10,000

one point crossover: 0.8

swap mutation: 0.004log2(M)

Unweighted X Weighted

	Fitness	Kate's Original	MCA	ECA	Effect Size
Unweighted	Cohesion	51	59.30 ± 1.10	59.37 ± 1.08	-
	Coupling	10	1.70 ± 1.10	1.63 ± 1.08	-
	Number of Clusters	2	2.57 ± 0.92	2.37 ± 0.87	-
	MQ	1.308	1.42 ± 0.28	1.33 ± 0.36	-
	Isolated Clusters	0	0.53 ± 0.76	-	-
	Cluster Difference	11	-	14.03 ± 7.79	-
Weighted	Cohesion	250	259.83 ± 4.62	258.73 ± 5.23	-
	Coupling	21	11.17 ± 4.62	12.27 ± 5.23	-
	Number of Clusters	2	5.90 ± 1.04	**6.97 ± 1.54**	0.22
	MQ	1.69	2.88 ± 0.46	2.71 ± 0.55	-
	Isolated Clusters	0	2.27 ± 1.26	-	-
	Cluster Difference	19	-	21.23 ± 2.03	-

Omnipresent Results

	Fitness	Kate's Original	MCA	ECA	Effect Size
$o_t = 3$	Cohesion	34	35.60 ± 1.36	35.47 ± 1.54	-
	Coupling	5	3.40 ± 1.36	3.53 ± 1.54	-
	Number of Clusters	2	5.07 ± 1.44	4.77 ± 1.69	-
	MQ	1.32	3.32 ± 1.02	3.11 ± 1.18	-
	Isolated Clusters	0	0.27 ± 0.44	-	-
	Cluster Difference	16	-	12.63 ± 4.03	-
$o_t = 2$	Cohesion	29	27.20 ± 0.95	27.67 ± 0.91	-
	Coupling	0	1.80 ± 0.95	1.33 ± 0.91	-
	Number of Clusters	2	**5.70 ± 1.04**	4.17 ± 1.75	0.73
	MQ	1.40	**3.96 ± 0.69**	2.93 ± 1.17	0.76
	Isolated Clusters	0	0.00 ± 0.00	-	-
	Cluster Difference	17	-	6.03 ± 2.99	-