Software Reuse Evaluation based on Functional Similarity in COSMIC-FFP Size Components

Luca Santillo (DPO, Italy)
Alain Abran (ETS, Canada)

SMEF, Rome, May 10-12, 2006

Agenda

- Size & Reuse
- Functional Similarity Assessment
  - 1st order evaluation
  - 2nd order evaluation
- Empirical Data Set
- Assessement Results
- Conclusions

Size & Reuse

- “Cost drivers”
  - Size – mainly
  - Reuse – significantly
- Determine potential for reuse
  - Find opportunities for reuse – assessement
  - Find them early – from functional requirements
- Exploit measurement elements

Functional Similarity – Context

- Software system as a hierarchy
  - Compare underlying structure of components
- Rate similarity
  - *Two functions are considered similar if they can be decomposed into the same subset of data movements and/or data manipulations.*
- Approximation orders
  - From functional processes to # Data Movements ("DMov's") and/or # Data Manipulations ("Dman's")
  - From human-based comparisons to side-by-side objective comparisons of components

1st Order Evaluation – Dmov's only

- % "same" DMov's across the functional processes being compared
  - F.P. "A": DMov's A1, A2, A3, A4, A5
  - F.P. "B": DMov's B1, B2, B3
  - If A1/B1, A2/B2, A3/B3 are "the same", then:
    - *A: 100% "similar" to B*
    - *B: 60% "similar" to A*
- Option: similarity matrix (discretized)

Shared DMs	Null (<10%)	Low (10-30%)	Avg (30-70%)	High (70-95%)	Max (>95%)
Similarity Value	0% (entirely different)	20%	50%	80%	100% (nearly identical)

2nd Order Eval. – DMov's & DMan's

- Orthogonal dimensions – independent

			Shared DMovs		
	Null (<10%)	**Low (10-30%)**	**Avg (30-70%)**	**High (70-95%)**	**Max (>95%)**
Shared DMans	-	-	-	-	-
Null (<10%)	0%	5%	10%	20%	40%
Low (10-30%)	5%	20%	30%	40%	50%
Avg (30-70%)	10%	30%	50%	60%	70%
High (70-95%)	20%	40%	60%	80%	90%
Max (>95%)	40%	50%	70%	90%	100%

- Note: COSMIC-FFP allows for local measurement extensions – proposal for data manipulation primitive actions (next)

Action-type list for DMan's

No.	Action	COSMIC-FFP Function Types
1	Data acceptance from outside the system's boundary	Data Movement (Entry-type, E)
2	Data presentation outside the system's boundary	Data Movement (eXit-type, X)
3	Data group reference/retrieval (read)	Data Movement (Read-type, R)
4	Data group insert/update (write)	Data Movement (Write-type, W)
5	Derived data creation by transforming existing data	Data Manipulation (creation, D)
6	Mathematical formulas/calculations	Data Manipulation (creation, M)
7	Condition analysis to determine which are applicable	Data Manipulation (check, A)
8	Data validation	Data Manipulation (check, V)
9	Equivalent-value conversion	Data Manipulation (check, C)
10	Data filtering/selection by specified criteria	Data Manipulation (check, F)

- Adapted from 1st gen. methods
- Some actions are already considered as DMov's (E/X/R/W)

Empirical Data Set

No.	Software System	Reference Document	Functional Processes	Size (Cfsu)	Verification Level
1	Automatic Line Switching (ALS)	ISO 14143-4 - RUR B8	14	66	C
2	Gateway Application (SAGA)	ISO 14143-4 - RUR B10	19	117	B
3	Valve Control (VC)	ISO 14143-4 - RUR B9	1	12	C
4	Hotel Reservation System (HRS)	ISO 14143-4 - RUR A1	7	66	C
5	L-Euchre System (LES)	ISO 14143-4 - RUR B11	15	61	B
6	Rice Cooker (RC)	Rice Cooker Requirements	3	12	D
7	Course Registration System (CRS)	CRS-RUP	19	96	C
8	Collegiate Sports Paging System (CSPS)	CSPS-RUP	27	136	B

- 2005 research on FSM standard “etalons”
- Verification level from A (min) to F (max)
  - verified by [measurer, indipendent expert, COSMIC leader, …, …, …, ISO IS]

Assessement Results – 1st Order

- 3 assessement criteria
  - "Same DMov" (DMov's that are the same share not only their own type and the underlying data group, but also the data portion that they actually move);
  - "Same DMov 'type'" (same type and same data group, but possibly slightly different subsets of data portions being moved);
  - [where the above criteria could not be applied] analyst's best judgment (apparent similarity of processes descriptions, of their triggering events, their data movements or their data groups).

Assessement Results – 1st Order

No.	Case Study ID.	Number of Functional Processes	Avg. Size per Functional Process (in Cfsu)	$MinSim_{fp}$	$MinSim_{avg}$	AvgSim	$MaxSim_{avg}$	$MaxSim_{fp}$
1	ALS	14	4.7	40%	52%	73%	100%	100%
2	SAGA	19	6.2	0%	0%	10%	27%	75%
3	VC	1	9.4	0%	0%	0%	0%	0%
4	HRS	7	12.0	0%	23%	61%	88%	100%
5	LES	15	4.1	0%	0%	8%	37%	67%
6	RC	3	4.0	0%	0%	9%	18%	33%
7	CRS	19	5.1	0%	20%	28%	68%	100%
8	CSPS	27	5.0	0%	0%	9%	45%	75%

- $MinSim_{fp}$ – minimum assessed similarity, per functional process
- $MinSim_{avg}$ – average of minimum values, over all functional processes
- AvgSim – average of assessed similarity, over all functional processes
- $MaxSim_{avg}$ – average of the maximum values over all functional processes
- $MaxSim_{fp}$ – maximum assessed similarity, per functional process

source

Assessement Results – 1st Order

- Case 1 with 73% (AvgSim) & Case 4 with 61% - high potential reuse
- Case 5 with 8% (AvgSim) & Case 8 with 9% - little potential reuse
- No specific similarity pattern related to the average size per functional process (e.g. cases 5 & 6, or 7 & 8)
- Comparison process – theoretically $N^2 – N = N(N - 1)$ comparisons. In practice – quicker
  - Comparison is transitive A vs. B = B vs. A
  - Filtering measurement elements helps accelerate comparisons
- Cases when both $MinSim_{fp}$ & $MinSim_{avg}$ equal 0% (taking *any* of the functional processes, there is at least one *other* functional process which has nothing in common with the that one)
  - The system can be divided into 2+ subsystems having “nothing in common”
  - Trivial example: login process (disjointed from any other process).

Assessement Results – 2nd Order

No.	Case Study I.D.	Number of Functional Processes	Avg. DMan's Per Functional Process (count)	$MinSim_{fp}$	$MinSim_{avg}$	AvgSim	$MaxSim_{avg}$	$MaxSim_{fp}$
1	ALS	14	1.0	10%	10%	49%	100%	100%
2	SAGA	19	0.9	0%	0%	5%	30%	70%
3	VC	1	5.0	0%	0%	0%	0%	0%
4	HRS	7	2.4	0%	6%	23%	51%	100%
5	LES	15	0.8	0%	0%	4%	28%	70%
6	RC	3	1.3	0%	3%	16%	28%	40%
7	CRS	19	5.1	0%	5%	8%	39%	90%
8	CSPS	27	5.0	0%	0%	3%	20%	70%

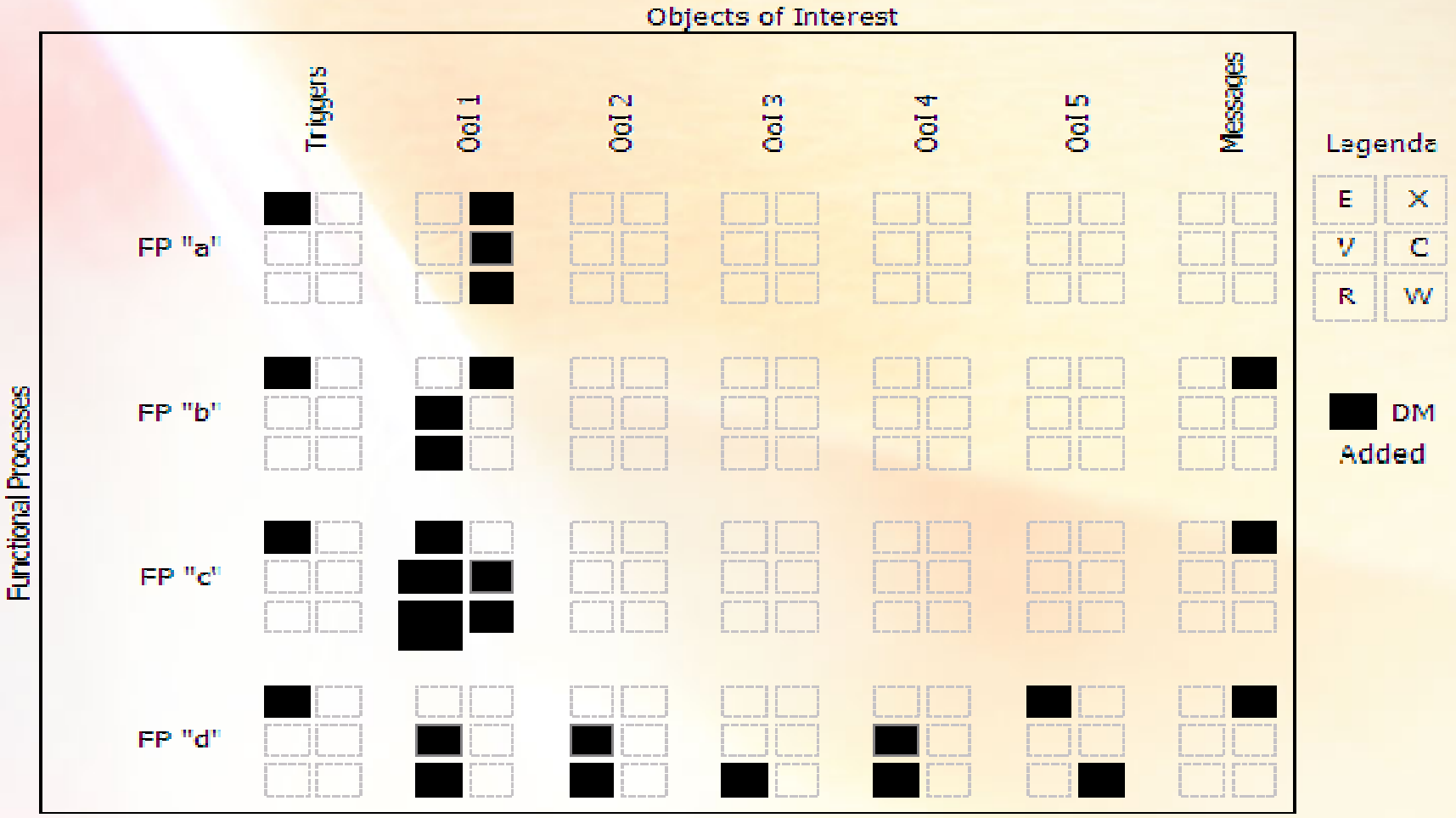
- $MinSim_{fp}$ – minimum assessed similarity, per functional process
- $MinSim_{avg}$ – average of minimum values, over all functional processes
- AvgSim – average of assessed similarity, over all functional processes
- $MaxSim_{avg}$ – average of the maximum values over all functional processes
- $MaxSim_{fp}$ – maximum assessed similarity, per functional process

source

Assessement Results – 2nd Order

- Case 1 w49% (AvgSim) & Case 4 w23% - still good potential reuse
- Several functional processes were found where no specific DMan's action was identified (e.g processes designed to simply "pass over" information between the system and its user by means of DMov's)
- Actually, not all averages diminished – as expected – from 1st to 2nd order evaluation (details depend on the proposed similarity matrix)
- Again, filtering speeds up the comparison process.
  - Several processes involved by different data groups are not to be compared (unless technical reuse is searched)

Visualization – Func.Proc.Diagram



Conclusions

- Relative ease of application of the comparison criteria
- The proposed technique is promising for real-world application (where functional measurement is applied)
  - The intent of the proposed approach is not to provide an exact number of candidates for reuse, but a reasonable assessment of that number (very useful to management for planning purposes).
  - A “precise” answer about reuse would require much more analysis time & effort.
- Further developments
  - Similarity evaluation for technical reuse
  - Refinement of similarity evaluation criteria
  - Refinement or extension of case studies

Thanks

- Luca Santillo
  - luca.santillo@dpo.it
- Alain Abran
  - aabran@ele.etsmtl.ca

Questions?