Extracting and Modeling Product Line Functional Requirements

Nan Niu, Steve Easterbrook

Department of Computer Science
University of Toronto, Canada
Email: {nn, sme}@cs.toronto.edu

Sept. 11, 2008

Product Lines Are Everywhere

Big Mac

Filet-O-Fish

McRib

Big N' Tasty

Core Assets: Benefit or Burden?

- ⇨Reuse is planned, enabled, and enforced
  - ⇨Core assets: design *for* reuse
  - ⇨Products: design *with* reuse
- ⇨Core assets ≠ actual product
  - ⇨Not worth building by itself
- ⇨Domain analysis
  - ⇨Heavyweight & no rules

costly → cannot afford

Extractive SPL Adoption Model

- [Krueger, 2001]
  - Reusing existing products for the SPL's initial baseline
  - SMEs (small and medium-sized enterprises)
  - Rather under-specify than over-specify
    - Reactive development can overcome under-specifying

Outline

⇨Problem

⇨Extracting

⇨Modeling

⇨Conclusions

Requirements Assets

- ⇨Principles [Pohl et al., 2005]
  - ⇨Focus more on *external* variability (visible to customers) and less on *internal* variability (useful to implementers)
  - ⇨Focus more on *what* varies (variation point) and less on *how* it varies (variants)

- ⇨Challenges of extraction
  - ⇨Where, what, & how to extract?
  - ⇨How to represent (model) the results?

Illustrative Example

3.1 Functions

...

3.1.3.2 Marking

AMS shall store students' assignments.

Instructor shall create the marking rubric.

Marker shall mark students' assignments.

AMS shall apply late policy automatically.

AMS can generate report for each section.

...

3.1.3.4 Remarking

Students can request remarking to markers.

AMS shall access marked assignments.

...

(a) SRS for AMS (Assignment Marking System)

3.1 Functional requirements

...

3.1.1 The EMS shall time–stamp any information sent online.

...

3.1.4 A professor must create a marking scheme for an assignment. A professor specifies mark breakdown, and records this information using a marking rubric.

...

3.1.8 A TA shall mark any portion of an assignment.

...

(b) SRS for EMS (Electronic Marking System)

- ⇨ Where to extract?
  - ⇨ Natural language documents

What to Extract?

- ⇨Functional requirements
  - ⇨Salient features directly observable by the users and other stakeholders
  - ⇨Basis for aligning and optimizing quality requirements

- ⇨Research question
  - ⇨Given a natural language document, how can its characterizing attributes, which relate to system functionalities, be produced?

Single-term Indices

- Verb_Freq [John, 2001]
  - Highest frequencies of occurrence

- Verb_INFO
  - $INFO(w) = -\log_2(P\{w\})$

Verbs	Freq.	Verbs	INFO
mark	64	accept	11.67
access	18	install	11.67
submit	14	highlight	11.67
use	9	update	10.67
notify	9	check	10.09
...	...	...	...

Multi-term Indices

- [Maarek et al., 1991]
  - Using LAs (lexical affinities) to profile software libraries
  - Information Retrieval based technique: cost, scalability, and domain transportability
  - Shortcoming #1: strictly two-word long
    - Domain-aware: "marking rubric" ➔ "marking_rubric"
  - Shortcoming #2: mixed modifier-modified relations
    - NL program analysis: "Verb-Direct Object" pairs can represent action-oriented concerns [Fry et al., 2008]

FRPs

- Functional Requirements Profiles
  - Action-oriented concerns that bear a high information value of a document
  - Model user-visible system functionalities
  - Represented by "verb-direct object" pairs

- Sample auto-marker FRPs
  - mark assignment, view grade, create marking rubric …

Profiling An Auto-Marker SRS

Verbs	Freq.	Verbs	INFO	FRPs	ρ
mark	64	accept	11.67	mark assignment	106.75
access	18	install	11.67	access assignment	96.31
submit	14	highlight	11.67	divide number	69.85
release	14	define	11.67	notify instructor	61.97
provide	12	calculate	10.67	modify information	54.68
use	9	update	10.67	release assignment	44.58
notify	9	check	10.09	change password	43.10
...	...	...	...	...	...

Quantity of information:

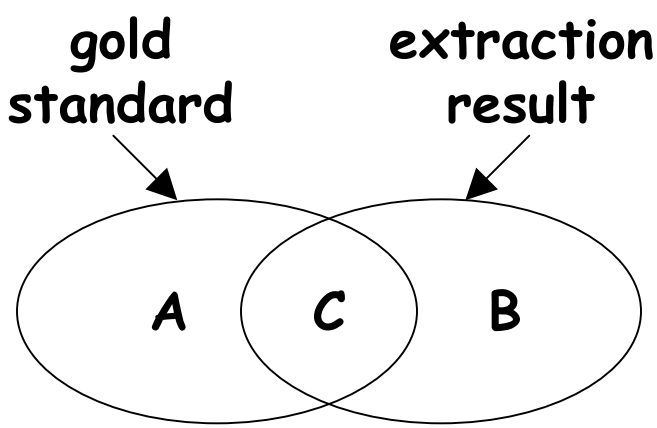
$$\begin{aligned}\rho((\{u_1, u_2\}, f)) &= f \times \mathrm{INFO}(\{u_1, u_2\}) \\ &= f \times -\log_2(P\{u_1, u_2\}) \\ &\approx f \times -\log_2(P\{u_1\} \times P\{u_2\})\end{aligned}$$

Effort

- Linear
  - Empirical observation: 98% of lexical affinities relate words which are separated by at most 5 words within a single sentence
  - 5-word window: a property of English

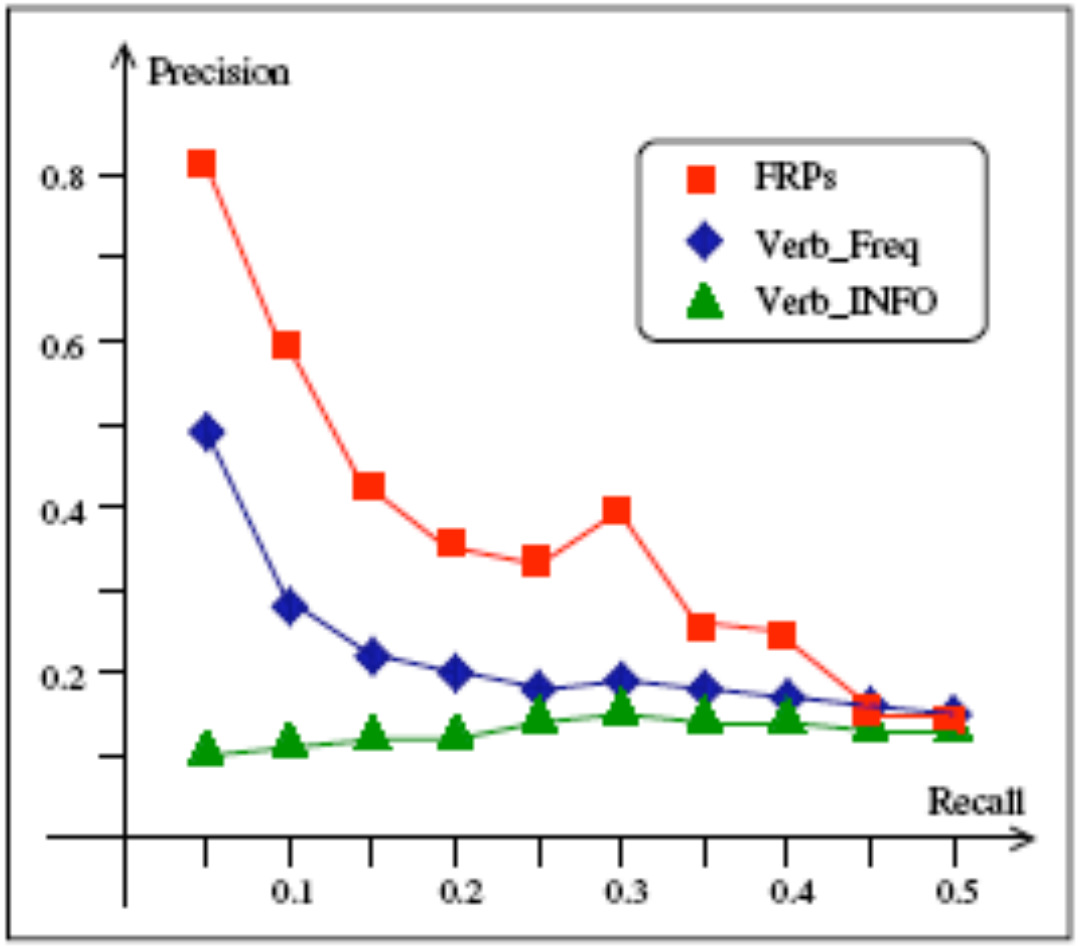
- OpenNLP
  - Stemming & POS tagging
  - Overhead is low

Effectiveness



Precision (accuracy) = C / B

Recall (coverage) = C / A



Outline

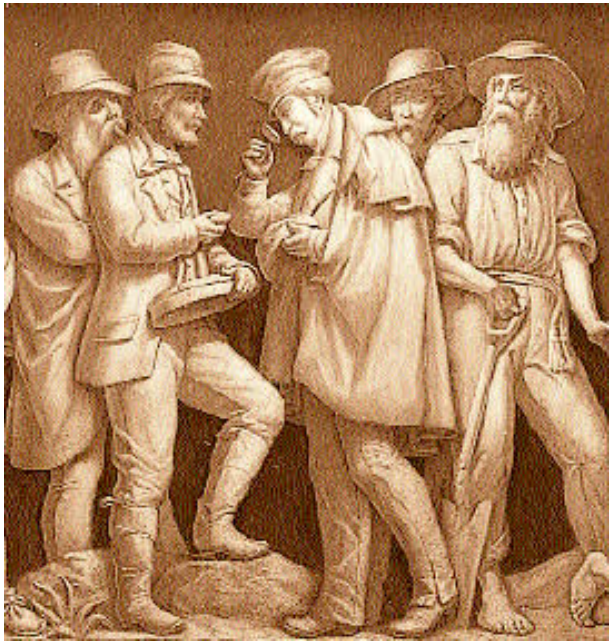
⇨Problem

⇨Extracting

⇨Modeling

⇨Conclusions

costly



Modeling FRPs

- ⇨Previous work
  - ⇨[Pohl et al., 2005]
    - ⇨OVM (orthogonal variability model): variation points, variants, dependencies, and constraints
    - ⇨Consistent variability management across requirements, design, realization, and testing
  - ⇨[Liaskos et al., 2006]
    - ⇨Using Fillmore's semantic cases to identify variability in goal models
    - ⇨Goal concern (e.g., "send message") ≈ FRP

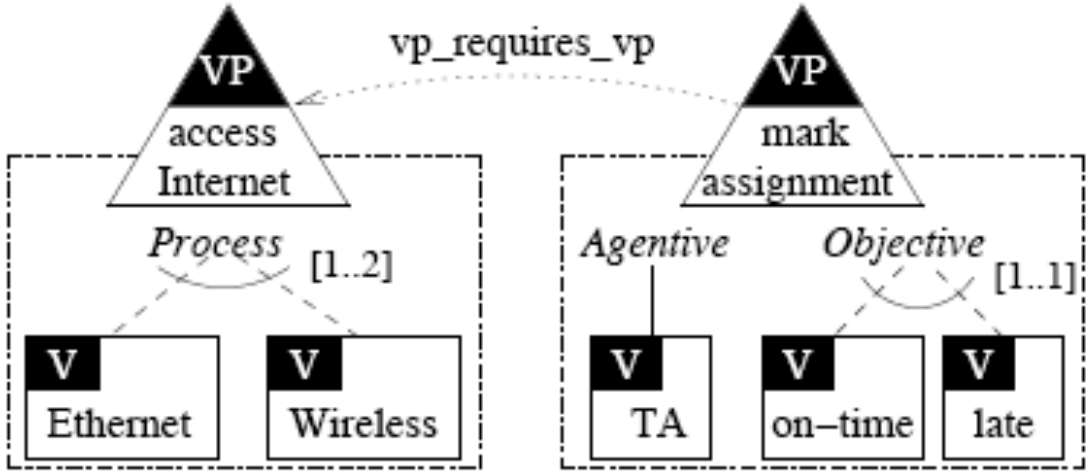
Uncovering Variation Structures

- Essential semantic cases (variation dimensions)
  - Agentive
  - Objective
  - Process
  - Conditional

FRP$_1$: mark assignment	
Agentive:	TA
Objective:	late, on–time
Process:	marking rubric, comment list, free form
Conditional:	access Internet
...	

FRP$_4$: access Internet	
Agentive:	student, instructor, TA, administrator
Process:	Ethernet, Wireless
...	

(a) Cases for FRPs



(b) Orthogonal variability model

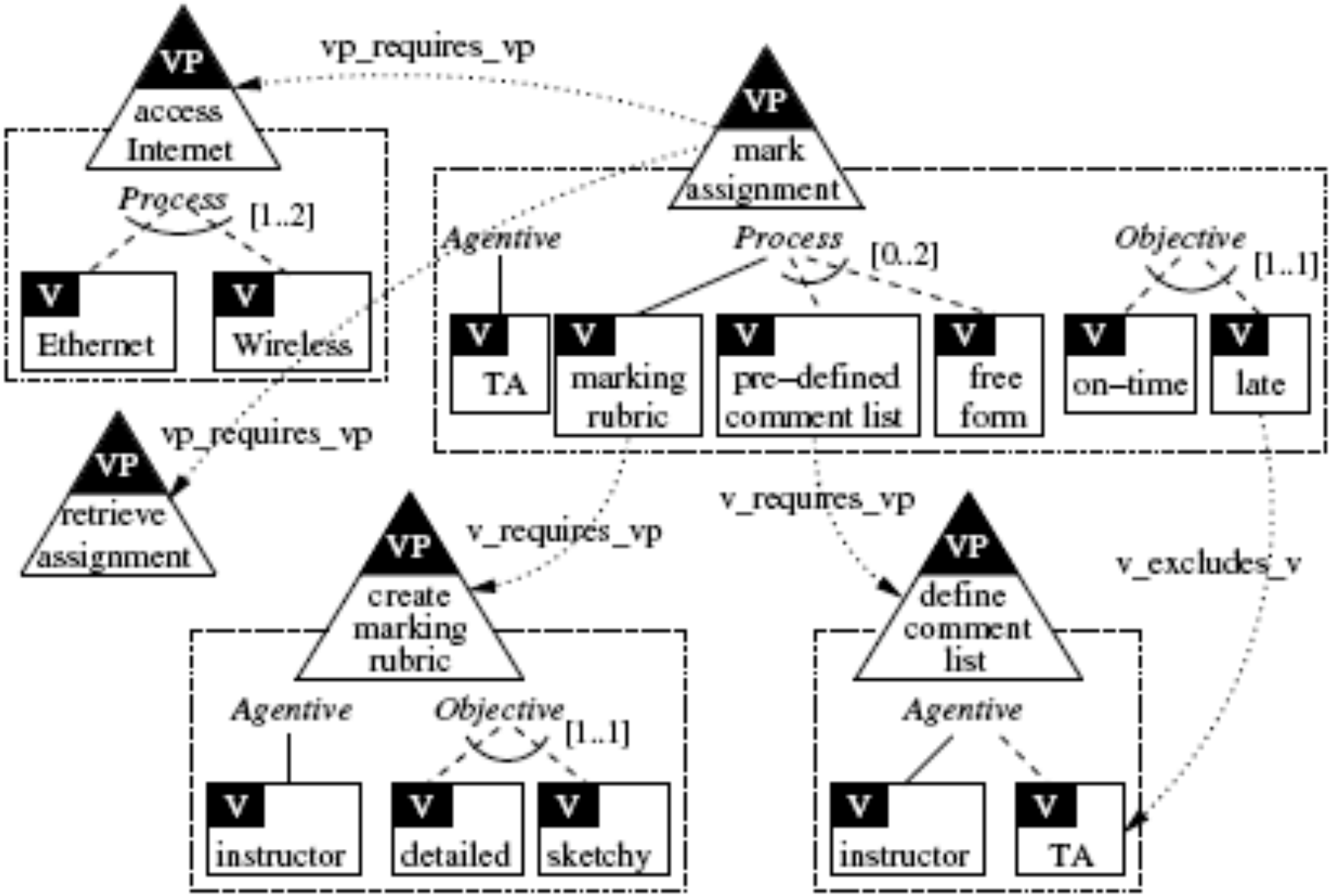
Semantic Case Analysis

- ⇨Focus more on WHAT and less on HOW
  - ⇨Reactively enriched
- ⇨ Treating each FRP as a variation point
  - ⇨Every product in the product line should address the FRPs (domain's action themes)
  - ⇨Other concerns can be variation points
- ⇨Manual effort is indispensable
  - ⇨NLP tools can help

Discovering Variability Dependencies

- Intra-FRP variability: mandatory or optional
  - Heuristic 1: If a case is associated with only one value, then the case has one mandatory variant.
  - Heuristic 2: (in the paper).

- Inter-FRP variability: requires or excludes
  - Heuristic 3: If FRP_2 is conditional to FRP_1, then there exists a vp_requires_vp constraint from FRP_1 to FRP_2.
  - Heuristic 4: (in the paper).


VP
access Internet
vp_requires_vp
VP
mark assignment
Process
[1..2]
V
Ethernet
V
Wireless
Agentive
Process
[0..2]
Objective
[1..1]
V
TA
V
marking rubric
V
pre–defined comment list
V
free form
V
on–time
V
late
VP
retrieve assignment
vp_requires_vp
VP
create marking rubric
v_requires_vp
v_requires_vp
VP
define comment list
v_excludes_v
Agentive
Objective
[1..1]
V
instructor
V
detailed
V
sketchy
Agentive
V
instructor
V
TA

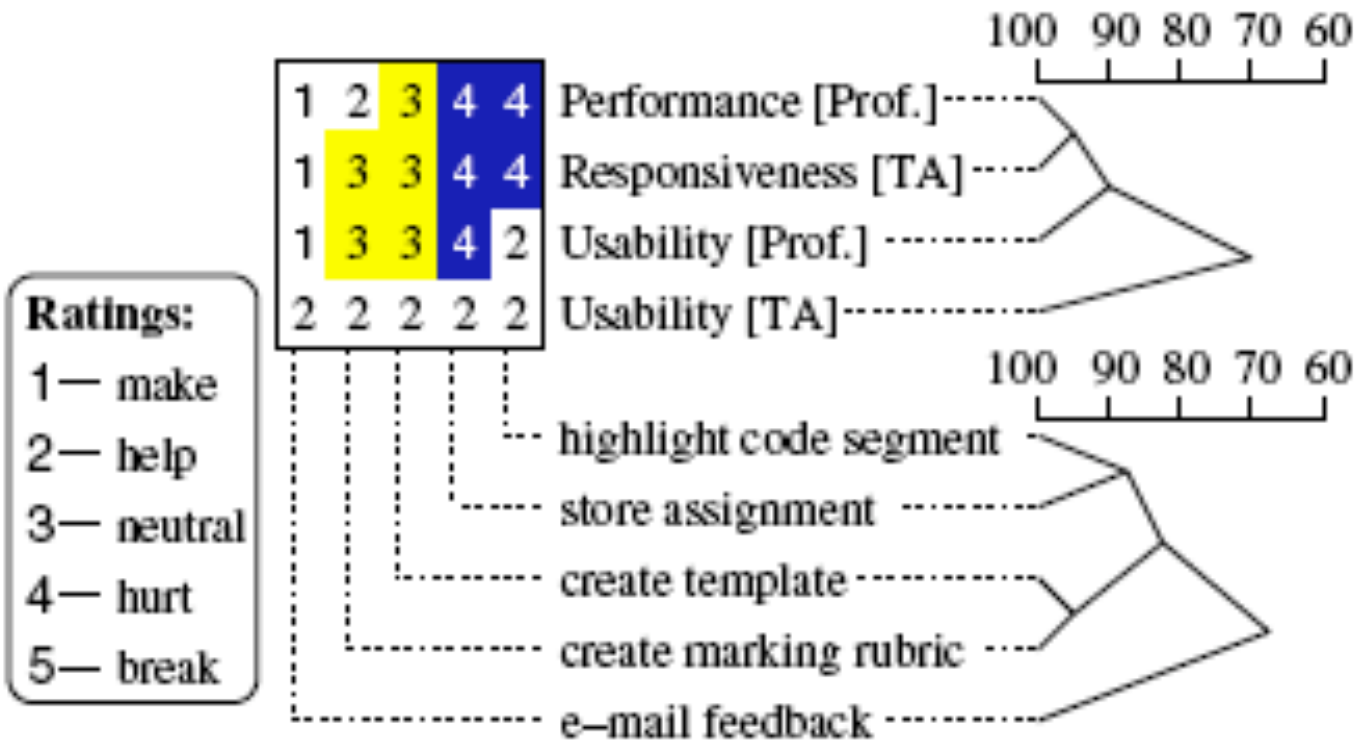
Outline

- ⇨Problem
- ⇨Extracting
- ⇨Modeling
- ⇨Conclusions

costly

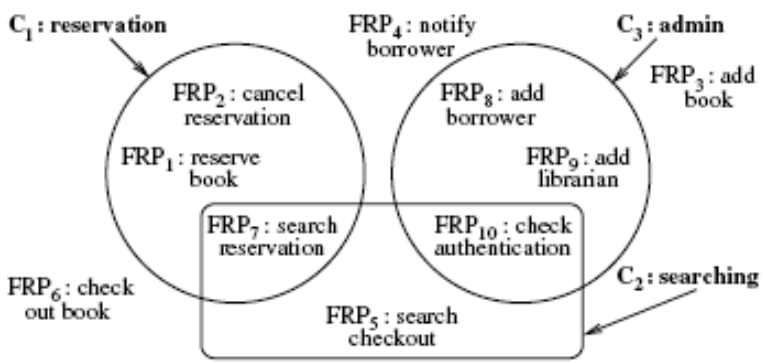
Aligning Quality Requirements

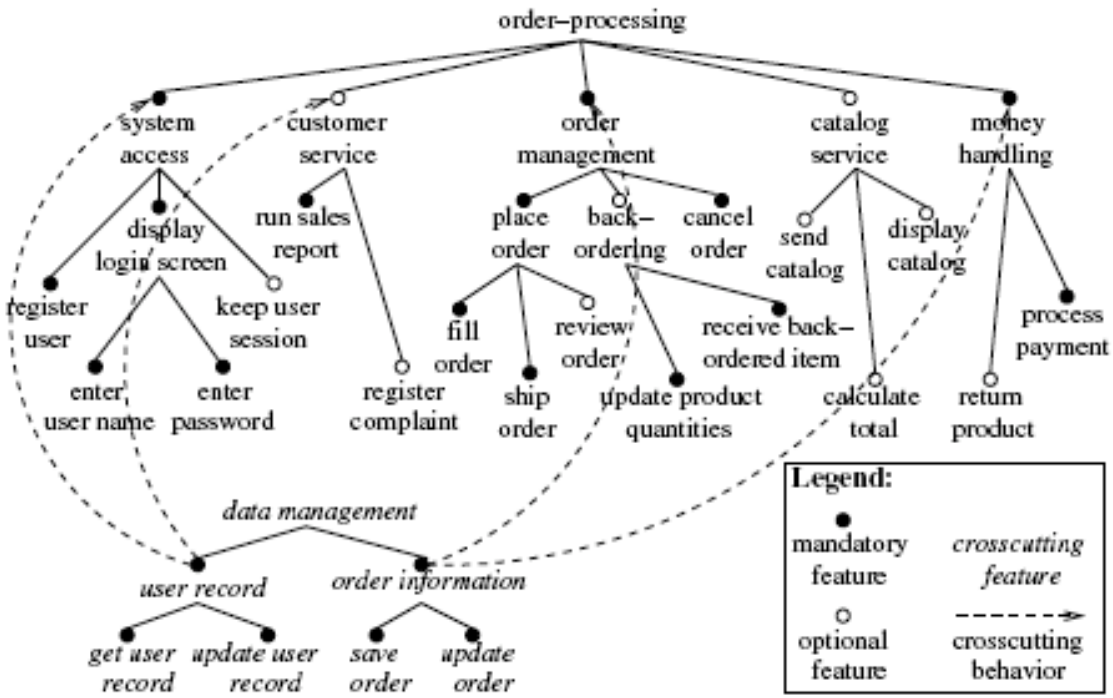
- [Niu & Easterbrook, RE 2006]
  - Tackling terminological interferences with the Repertory Grid Technique
  - FRPs: common ground



Cluster Analysis

- ⇨[Niu & Easterbrook, SPLC 2008]
  - ⇨User: identify, browse, prioritize features
  - ⇨Designer: system decomposition and modularization





Concluding Remarks

⇨FRPs

Acknowledgments

- CSC340 (Winter 2007) students and staff at University of Toronto
- Jennifer Campbell, Jia Wang, Sotirios Liaskos
- Anonymous RE'08 reviewers