

Advanced topics in SE in Novi Sad PhD student's report

Miloš Savić

Department of Mathematics and Informatics
Faculty of Sciences
University of Novi Sad
Serbia



Content

- Overview of activities
- Activities related to SNEIPL
- Activities related to students' projects
- Conclusion

Content

- **Overview of activities**
- Activities related to SNEIPL
- Activities related to students' projects
- Conclusion

11/14/2013

3

Overview of activities

- **SNEIPL**
 - Identification of new fields of research and practice where SNEIPL can be useful
 - The base for future students' projects
- **Assistance in students' projects**
 - Two dimensional extensibility of SSQSA
 - eCFG → eCST transformation

11/14/2013

4

Content

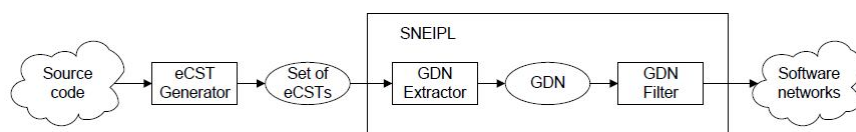
- Overview of activities
- **Activities related to SNEIPL**
- Activities related to students' projects
- Conclusion

11/14/2013

5

SNEIPL

- eCST-based extractor of software networks that provide different granularity views to software system:
 - Package collaboration networks
 - Class/module collaboration networks (CCN)
 - CCN restricted to specified coupling type (inheritance tree, aggregation network, etc.)
 - Static call graphs
 - FUGV graphs

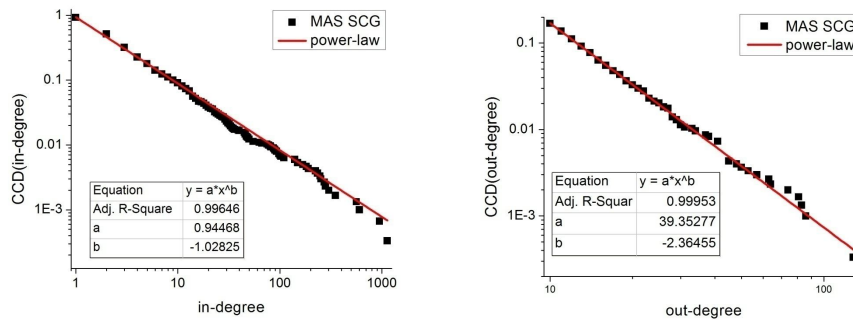


11/14/2013

6

SNEIPL, primary motivation

- Analysis of software systems under the framework of the complex network theory



Scale-free property of Modula-2 Algebra System

11/14/2013

7

SNEIPL, applicability in other fields

SNEIPL: eCSTs → GDN → software networks

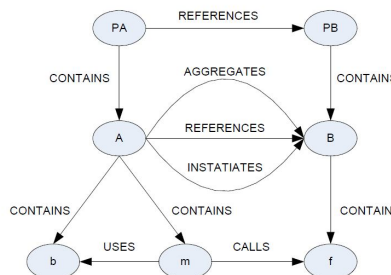
Source code:

```

package PA;
class A {
  B b = new B();

  void m() {
    b.f();
  }
}

package PB;
class B {
  void f() {
  }
}
  
```



- Metrics related to software design that quantify **invocation**, **coupling**, **cohesion** and **inheritance** among entities
- GDNs can be viewed as **fact-bases for reverse engineering tasks**

11/14/2013

8

Metrics with SNEIPL

- **Package-collaboration network**
 - Cecil Martin's afferent and efferent coupling
- **Class-collaboration network**
 - Chidamber-Kemerer (CK) coupling between objects (CBO)
- **Inheritance forest**
 - CK Number of Children (NOC) and Depth in inheritance tree (DIT)
- **Aggregation network**
 - QMOOD's measure of aggregation (MOA)
- **Hierarchy tree**
 - NOC for packages, Number of Methods/Attributes in a class
- **Static call graph + hierarchy tree**
 - CK Response for a class (RFC)
- **FUGV graph**
 - CK Lack of cohesion in methods (LCOM) and other cohesion metrics

11/14/2013

9

SNEIPL as fact-base extractor

- **Rigi, Moose, Gupro**
 - Language-independent representation of fact-bases, but **language-dependent** extraction of fact bases

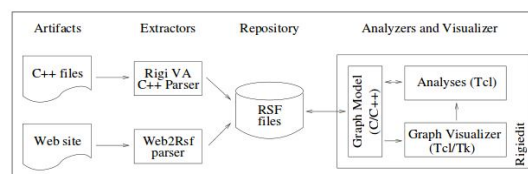


Fig. 1. Rigi's conceptual architecture

- **Bauhaus**
 - **Partially language-independent** extraction of fact bases
 - C, C++: automatic IML (InterMediate Language) → RFG (Resource Flow Graph)
 - Java, Ada: language-dependent RFG construction (class files/Ada Semantic Interface Specification)
- **Sneipl**
 - **Both language-independent extraction and representation of fact-bases**

11/14/2013

10

New SNEIPL export features

- SNEIPL currently exports GDN and other networks in txt and Pajek file formats
- Export of GDN in XML formats enables the integration of SNEIPL with
 - Reverse engineering tools
 - GXL (Graph eXchange Language) XML schema
 - Network visualization tools
 - GraphML XML schema
- The project currently assigned to a PhD student and still in progress

11/14/2013

11

Content

- Overview of activities
- Activities related to SNEIPL
- **Activities related to students' projects**
- Conclusion

11/14/2013

12

Overview of students' projects

- Two dimensional extensibility of SSQSA
 - **Task:** Extend SSQSA with a new language (Delphi) and new metrics (Halstead metric set)
 - **Team:** One undergraduate student from Novi Sad, Serbia
 - **My role:** to supervise the incorporation of eCST universal nodes related to Halstead metrics in tree-rewrite rules of the grammar and the implementation of the algorithm that computes Halstead metrics
- eCST → eCFG transformation
 - **Task:** Extend eCST with control-flow links
 - **Team:** Two master students from Skopje, Macedonia
 - **My role:** To design and supervise the complete project

11/14/2013

13

Two dimensional extensibility of SSQSA

- **Communication:** SSQSA/ATSE meetings, e-mail consultations
- **The problem with Halstead metrics is that it is not clearly defined what is counted as Halstead's operator and what as Halstead's operand.**
- Student instructed how to adopt the following language-independent schema:
 - **Halstead's operators:** tokens introduced by language designers (keywords, separators, operators)
 - **Halstead's operands:** tokens introduced by software developers (identifiers, constants)
- Before started to work on the program that computes Halstead metrics the student sent me the grammar for a revision.
 - So, the supervisor was in the position to act proactively

11/14/2013

14

The first revision of the grammar

- Hierarchical constraints were not completely satisfied
 - Lexical-level universal nodes mark tokens, not syntactic constructions (cannot be roots of non-trivial eCST sub-trees)
 - The constraint that is directly related to the computation of Halstead metrics: enables identification of unique Halstead's operators and operands
- Grammar completely finished in two iterations before the implementation of the program
 - Prevented dilemma: is the problem in the grammar or in the program?

11/14/2013

15

Why the project was successful?

- Some of possible explanations:
 - Motivated student
 - Student highly interested in compiler construction and intermediate representations of source code
 - Technical skills
 - Student familiar with various parser generators
- **Fact:** Continuous work on the project combined with the communication initiated by the student

11/14/2013

16

eCFG → eCST project

- eCFG = extended Control-Flow Graph
- eCFG = eCST + control-flow links
- Control-flow links connect statements in procedures/methods, i.e. STATEMENT universal nodes

11/14/2013

17

Project description document

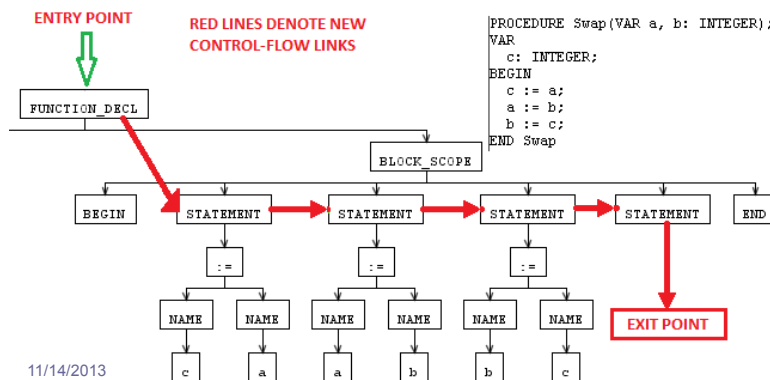
- Students not from Novi Sad, comprehensive introduction to the project
- Document that contains
 - Basic definitions
 - Elaboration of the idea and the proof of concept
 - Warm-up exercises
 - Technical guidelines
 - Illustrated examples
 - Project success criteria
 - Relevant literature

11/14/2013

18

Proof of the concept

- Explains how to add control-flow links in
 - Simple linear sequence of statements
 - Single and multiple branch statements
 - Loop statements (pre-condition, post-condition and fixed loops)
 - Jump statements



Communication with students, timeline

- 27.3.2013 – project description document sent to students
- 19.5.2013 – reminder to live consultations in Skopje, request for a progress report, e-mail consultations
- 29.5.2013 – live consultations in Skopje, demonstration of what students made

Observed problems

- No response after the project description document was sent
- Not careful reading of the project description document

... we are sorry to be late, but **we had trouble with AntlrWorks**. First we've downloaded **AntlrWorks4** and had problem with the new grammar. We tried to change the SSQSA_Java.g into AntlrWorks4 grammar, but no success there. After that we downloaded **Version 1.5**, and still we can't get it to draw the eCST tree.

Excerpt from the project description document:

Example 2: The results of Task 5 for linear search written in Java (**AntlrWorks 1.4.3 is used**). At the left side is the source code, while at the right side resulting eCST tree is shown.

- Live consultations:
 - Project description document is a poor substitute for live meetings
 - one active and one passive team member during the demonstration → **team work?**
 - Students clearly understood the problem, know how to solve it, but showed no interests in the background of the project and its potential contributions → **are they motivated to work on the project?**

11/14/2013

21

Content

- Overview of activities
- Activities related to SNEIPL
- Activities related to students' projects
- **Conclusion**

11/14/2013

22

Conclusions

- “Learning through research” paradigm works with highly motivated students
- **Motivation is a big problem**
 - Undergraduate and master students
 - Usually not interested in the importance of a project, its background and potential contributions
 - PhD students
 - Interested in a bigger picture (critical thinkers), but projects are not closely related to their fields of research
 - “Is this project related to native XML databases?”

11/14/2013

23

Advanced topics in SE in Novi Sad PhD student's report

Miloš Savić

Department of Mathematics and Informatics
Faculty of Sciences
University of Novi Sad
Serbia