

# CS3215 (2011): Project Handbook

Stan Jarzabek, January 2011

## Table of contents:

1	An overview of the project handbook and project resources.....	2
2	How to succeed in the SE Project Course?.....	3
2.1	Project in glance.....	4
3	The Policy on Project Work .....	4
4	A brief problem description.....	6
4.1	Motivation.....	6
4.2	What is an SPA and how is it used?.....	6
4.3	How does an SPA work? .....	6
5	Source language <i>SIMPLE</i> .....	7
	Other rules: .....	9
6	Program design abstractions for <i>SIMPLE</i> .....	9
6.1	A model of global program design abstractions.....	9
6.2	A model of abstract syntax of <i>SIMPLE</i> .....	11
6.3	A model of program control and data flow .....	14
6.4	Summary of program design models .....	16
7	Querying programs with <i>PQL</i> .....	17
7.1	General rules for writing program queries in <i>PQL</i> .....	17
7.2	Types of relationship arguments in program queries .....	18
7.3	Query examples: querying global program design information.....	19
7.4	Query examples: querying control and data flow information .....	20
7.5	Query examples: finding syntactic code patterns.....	20
7.6	Rules for patterns .....	22
7.6.1	Well-formed pattern specifications.....	22
7.6.2	Matching patterns .....	22
7.6.3	Using an underscore.....	22
7.7	Comments on query semantics.....	22
7.7.1	Implicit existential quantifier in query conditions .....	22
7.7.2	Implicit 'and' operator between query clauses .....	22
7.7.3	Use of free variables .....	22
7.7.4	A note on meaningless queries .....	23
7.8	A general format for <i>PQL</i> queries.....	23
8	Required program quality attributes .....	24
8.1	Flexibility of the SPA .....	24
8.2	Reusability .....	24
8.3	Reliability of your programs and testing.....	24
8.4	Scalability.....	25
8.5	Quality of program documentation and project reports .....	25
8.6	Performance of query evaluation .....	25
9	An architecture of an SPA - towards the program solution .....	26
9.1	What is a software architecture? .....	26
9.2	SPA architecture: component view .....	26
9.3	An SPA front-end subsystem.....	27
9.4	A PKB subsystem .....	27
9.5	Abstract versus Concrete PKB API .....	29
9.5.1	Benefits of abstract PKB API .....	29
9.5.2	How to use abstract PKB API in the project.....	29
9.6	How to discover abstract PKB API.....	30
9.7	How to document abstract PKB API .....	32
9.8	Concrete PKB API (class interfaces in C++ program) .....	33

10	Compendium of engineering practices for the project.....	33
10.1	Preliminaries: common sense software engineering practices .....	34
	Rule 1: Understand a problem before you go for full-fledged implementation.....	34
	Rule 2: Separation of concerns .....	34
	Rule 3: The virtues of simplicity and standardization .....	35
	Rule 4: System decomposition with information hiding .....	35
10.2	Becoming a great architect: Making right design decisions .....	35
10.2.1	General approach to making design decisions .....	36
10.2.2	Architectural design decisions in SPA .....	36
10.2.3	Detailed design decisions in SPA.....	36
10.2.4	Estimating algorithmic complexity: Big O notation .....	37
10.2.5	Documenting architectural design decisions .....	37
10.2.6	Documenting detailed design decisions .....	37
10.3	Using UML in the project.....	38
10.3.1	Using UML sequence diagrams .....	38
10.3.2	When to use sequence diagrams.....	40
10.3.3	Summary of recommendations: .....	40
10.4	Design patterns.....	41
10.5	Table-driven technique - query validation example .....	41
10.6	Error handling, exceptions and assertions in C++.....	42
10.6.1	Exceptions .....	43
10.6.2	Assertions.....	43
10.7	Testing your programs.....	44
10.7.1	You should do the following types of testing: .....	44
10.7.2	Test plans and test case documentation.....	45
10.8	Notes on documenting your programs.....	46
11	The project team .....	47
12	An SDLC for the project.....	47
12.1	Analysis .....	48
12.2	Architectural design.....	48
12.3	Incremental development .....	48
12.4	Planning incremental development .....	49
13	Technical tips.....	51
13.1	User interface to SPA .....	51
13.2	Parsing <i>SIMPLE</i> .....	51
13.3	Notes on AST .....	52
13.4	Generating an AST during parsing.....	52
13.5	Creating symbol tables during parsing .....	53
13.6	What information should be stored in the PKB?.....	53
13.7	Relationships Modifies and Uses .....	53
13.8	Processing <i>PQL</i> queries.....	53
13.8.1	Entering program queries (Group- <i>PQL</i> ) .....	54
13.8.2	Query pre-processor .....	54
13.8.3	Query evaluation and optimization .....	54
	References.....	56
	Appendix A. Summary of <i>PQL</i> grammar rules.....	56
	Summary of other <i>PQL</i> rules:.....	57

## 1 An overview of the project handbook and project resources

This handbook is your main text for the SE project course. The handbook explains in detail the problem you will be working on and methods you should use to meet project requirements. In sections 2 and 3, we emphasize the main points, for your reference throughout the course.

The project handbook consists of four parts:

**Part I: Software Requirements:** We explain functional requirements and quality attributes for the software system you will implement.

**Part II: Software Methods:** Here you will find guidelines for the methods you should use in the project.

**Part III: Software Process:** In this part of the handbook, look for everything that has to do with organizing a project work and a Software Development Life Cycle (SDLC) for the project.

**Part IV: Technical Tips:** Here you will find detailed technical tips that should help you solve specific design and implementation problems. Using examples, we show you how to implement a simple parser, generate an abstract syntax tree during parsing, apply table-driven technique, etc.

**References and appendix** with definitions of *PQL* syntax rules complete the document.

You should review the whole project handbook at least twice during the first weeks of the course (by the end of the Assignment 1), so that you understand the overall goal of the project.

Throughout the course, you should study in detail sections relevant to specific assignments, so that you are clear what methods we expect you to use during analysis, design, implementation, testing, documentation and report writing.

- You will find links to all the project resources at the course Web site.

## 2 How to succeed in the SE Project Course?

First of all, you must keep in mind the objectives for the project course which are the following:

- prepare students for industrial projects (fuzzy requirements, incomplete specifications, great deal of information to prioritize and ponder on),
- develop ability to work in group in a project of substantial size and complexity,
- enhance development skills in areas of design, construction, testing and documentation,
- enhance project planning skills,
- develop communication and writing skills,
- apply and consolidate what students have learned in the following three programming courses CS1101, CS1102 and CS2103,
- follow the Software Development Life Cycle according to the “best software engineering practices”.

In the project, you develop well-tested, almost production quality software system. In team projects, you deal with large volume of information. Not to get lost, you must understand what’s most important and stick to it. Here are some tips to help you focus on essentials and not to “lose forest for trees”:

1. Review [Required program quality attributes](#) and [Compendium of recommended engineering practices for the project](#) whenever you embark on a new project task.
2. Attitude “Just get a program run!” will not work in this course. Take your time to design, test and document your programs.
  - a) Your project will be evaluated based on the quality of program design, implementation and documentation. Not only functions implemented but also program reliability achieved through testing will count. Ad hoc development will not yield the qualities we will be looking for in your program solution and final report. Do not go for implementing much functionality if you cannot assure quality.
  - b) As you will develop a program incrementally, whatever you produce today, you will have to use and extend tomorrow. Ad hoc, poorly organized and documented program will cost you extra effort tomorrow. Eventually, it may lead to failure. Refine your documents to keep them up to date as you go through iterations.

- c) Focus on reuseability (develop once, reuse somewhere else), extensibility (design for change), and robustness (report an error if parts of the input are wrong). These are key factors in successful software development and will help you to easily evolve your program over time and adapt to changing requirements.
  - d) You work in a team. Work products (programs and documentation) you deliver are used by your teammates. To minimize miscommunication (you cannot eliminate it totally!), your work products must be easily understandable to others.
  - e) Interfaces among team members and among program components developed by different team members must be very well thought out and clear. Defining interfaces is “the heart” of architectural design, it is the key to the success of your project. Spend much time to get interfaces right - it will save your time as you progress through the project and will protect you against failure. Focus your weekly team meetings on discussing interfaces. We shall insist that you master the skill of defining interfaces and communicate with other teammates in terms of interfaces. This is one of the most important skills every good software designer must have.
3. Arrange team meetings on weekly basis. Have a clear agenda for each meeting. List problems you want to discuss and be clear what you want to achieve at the end of the meeting.
  4. The SE project Course is worth 8 modular credits and we expect you to spend time equivalent to two courses on the project work. Time management is a critical success factor for this course as you will need to juggle between this project and other coursework deadlines. Often, you will find yourself putting off this project in the face of other more urgent assignments’ deadlines. However, we would like to remind you that the success of this project relies on the continuous and consistent effort of *every* team member *throughout* the semester. Last minute work will only get you a poor grade. Hence, please weigh your priorities and manage your time accordingly.

We make it clear: writing just any program that works is not what we expect from you. We want you to develop well designed, documented and tested program, by following an incremental development process and by applying state of the art software methods.

## 2.1 Project in glance

Here is a snapshot of what you will be doing in the project course. You will start by studying the Project Handbook to understand the project objectives, problem you will be solving and software methods to be used. There will be lectures during first 3-4 weeks of the term to guide you through the project material. Each team will be assigned a slot for consultation with the supervisor.

The project phases are defined in five assignments. After analysis, architectural design and prototyping, you will develop your program solution incrementally, in three iterations. At the end of each assignment, you will submit a project report and present your solution to a supervisor.

## 3 The Policy on Project Work

We would like to inform you in advance of our expectations regarding academic conduct. Please ponder over the issues carefully so that you will not lose perspective in case you are tempted to violate the policy during the busy schedule of meeting your assignment deadlines.

For homework assignments and project assignments, you are permitted to discuss problem requirements and background material with anyone. However, the actual solutions your team hand in should be your own team work. Throughout the module, you may discuss background material, problem requirements, approaches to problem solutions, and design with anyone, but you may not view any code and document written for CS3215 by anyone not in your team, including past students. Furthermore, you may not reveal your code and document to any

students not in your team. The actual coding and documentation should be the work of your team only.

In the past, many academic misconduct cases have come about because of poor judgment on the part of students who feel that they are incapable of completing an assignment. This feeling usually arises when you find yourself temporarily tired, stressed, or desperate. Please bear in mind that the long-term consequences of an academic misconduct case will do much more damage to your career than the worst possible grade you can get in the course.

The course staff will not be happy to deal with any academic misconduct case, as this requires an enormous amount of non-productive effort, and involves numerous university resources. However, for the long-term benefit and fairness of all students, we will prosecute every such case to the fullest extent provided for by the University regulation.

# --- Part 1: Software Requirements ---

## 4 A brief problem description

### 4.1 Motivation

Some companies spend as much as 80% of software budgets on software maintenance. During software maintenance, programmers spend almost 50% trying to understand a program. Therefore, methods and tools that can ease program understanding have a potential to substantially cut computing costs.

During program maintenance, programmers often try to locate code relevant to the maintenance task in hand. Here are examples of questions programmers might ask to locate code of interest:

- Q 1. Which procedures are called from procedure “Second”?
- Q 2. Which procedures call procedure “Second”?
- Q 3. Which variables have their values modified in procedure “Second”?
- Q 4. Find assignment statements where variable  $x$  appears on the LHS.
- Q 5. Find statements that contain sub-expression  $x*y+z$ .
- Q 6. Find three while-do loops nested one in another.
- Q 7. Find all program statements that modify variable  $x$  and use variable  $y$  at the same time.
- Q 8. Is there a control path from statement #20 to statement #620?
- Q 9. Which assignments that modify variable  $x$  affect value of  $x$  at statement #120?
- Q 10. Which program statements affect value of  $x$  at statement #120?
- Q 11. If I change value of  $x$  in statement #20, which other statements will be affected?
- Q 12. Find all assignments to variable  $x$  such that the value of  $x$  is subsequently re-assigned a value in an assignment nested inside two loops.

To answer the above questions, a programmer may need to examine huge amount of code. Doing this by hand may be time consuming and error prone.

### 4.2 What is an SPA and how is it used?

A **Static Program Analyzer (SPA for short)** is an interactive tool that automatically answers queries about programs. Of course, an SPA cannot answer all possible program queries. But the class of program queries that can be answered automatically is wide enough to make an SPA an useful tool. In your programming practice, you might have used a cross-referencing tool. An SPA is an enhanced version of a cross-referencing tool. In this project, you will design and implement an SPA for a simple source language.

The following scenario describes how an SPA is used by programmers:

1. John, a programmer, is given a task to fix an error in a program.
2. John feeds the program into SPA for automated analysis. The SPA parses a program into the internal representation stored in a Program Knowledge Base (PKB).
3. Now, John can start using SPA to help him find program statements that cause the crash. John repeatedly enters queries to the SPA. The SPA evaluates queries and displays results. John analyses query results and examines related sections of the program trying to locate the source of the error.
4. John finds program statement(s) responsible for an error. Now he is ready to modify the program to fix the error. Before that, John can ask the SPA more queries to examine a possible unwanted ripple effect of changes he intends to do.

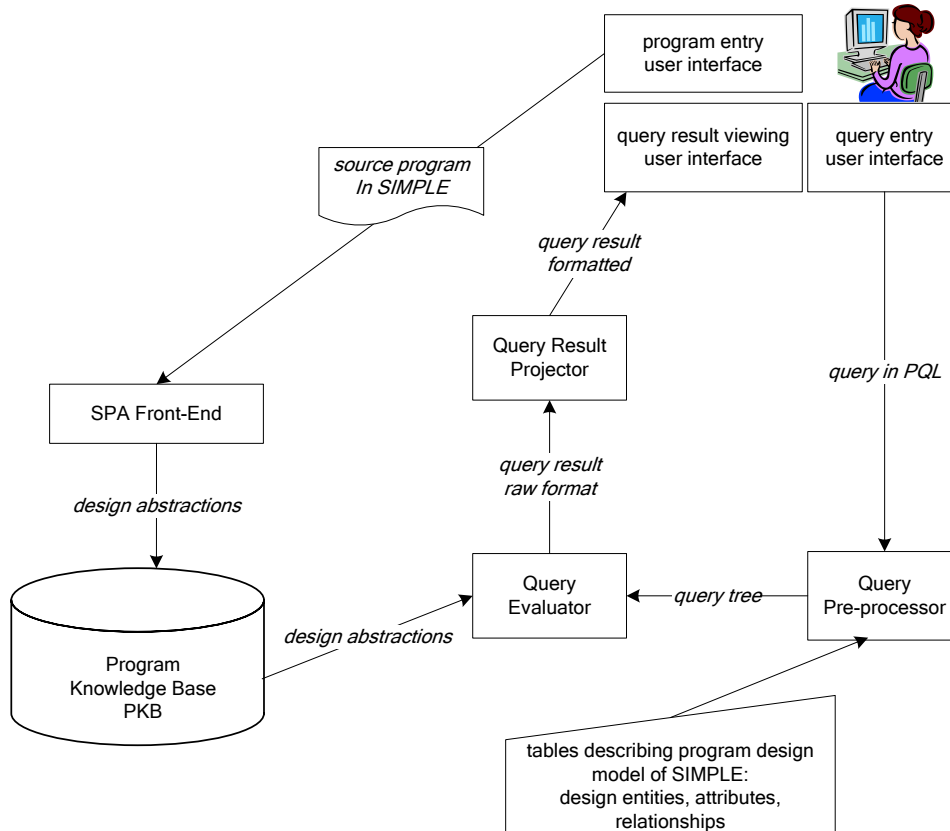
From a programmer’s point of view, there are three use cases: source program entering and automated analysis, query entering and processing, and viewing query results.

### 4.3 How does an SPA work?

In order to answer program queries, an SPA must first analyze a source program and extract relevant program design abstractions. Program design abstractions that are useful in answering queries typically include an abstract program syntax tree, program control flow graph and cross-reference lists indicating usage of program variables, procedure invocations, etc. An SPA

*front-end* (Figure 1) parses a source program, extracts program design abstractions and stores them in a Program Knowledge Base (PKB).

Now, we need to provide a programmer with means to ask questions about programs. While using a plain English would be simple for programmer, it would be very difficult for SPA. Therefore, as a workable compromise, we define a semi-formal program query language (*PQL* for short) for a programmer to formulate program queries. A *query pre-processor* validates a query and builds query internal representation suitable for evaluation (so-called query tree). *Query evaluator* answers queries and *query result projector* displays query results for the programmer to view.



**Figure 1. Logical components and operational scenario of a Static Program Analyzer.**

It should be easy for a programmer to write program queries. In particular, a programmer should be able to concentrate on the program itself rather than on the representation of program design abstractions in the PKB. For that reason, our *PQL* will allow a programmer to ask program queries in terms of *program design models* rather than in terms of *physical program representations* stored in the PKB. *PQL* is similar to SQL, but it is a higher level language than SQL. For example, query *Which procedures are called from procedure "Second"?* can be written in *PQL* as follows:

Q1. procedure p;  
**Select p such that** Calls ("Second", p)

To answer queries, *query evaluator* (Figure 1) will map program design entities referenced in a query (such as *procedure*) and their relationships (such as *Calls*) into the program design abstractions stored in the PKB in the "raw form". Next, query evaluator will interpret query conditions in **such that**, **with** and other query clauses to extract from the PKB program query results, i.e., program design entities that match the conditions.

## 5 Source language *SIMPLE*

Your SPA will analyze programs written in a source language called *SIMPLE*. *SIMPLE* was designed for the purpose of experimenting with SPA techniques, not as a language for solving

real programming problems. The language contains the minimum number of constructs to serve that purpose. The example below depicts a program structure and language constructs in *SIMPLE*. Program lines in procedure Second are numbered for ease of referencing in examples.

```

procedure First {
  x = 2;
  z = 3;
  call Second; }
procedure Second {
1. x = 0;
2. i = 5;
3. while i {
4.     x = x + 2 * y;
5.     call Third;
6.     i = i - 1; }
7. if x then {
8.     x = x + 1; }
   else {
9.     z = 1; }
10. z = z + x + i;
11. y = z + 2;
12. x = x * y + z;}
procedure Third {
  z = 5;
  v = z; }

```

**Figure 2. Sample program in *SIMPLE***

Here is a summary of the language rules for *SIMPLE*: A program consists of one or more procedures. Program execution starts by calling the first procedure in a program. By convention, the name of the first procedure is also the name of the whole program. Procedures have no parameters, cannot be nested in each other or called recursively. A procedure contains a body of statements. Variables have unique names and global scope. Variables can be introduced in a program at any time. All the variables are of integer type, require no declarations and are initialized to 0. There are four types of statements, namely procedure call, assignment, while loop and if-then-else. Decisions in while and if statements are made based on the value of a control variable: TRUE for values different than 0 and FALSE for 0.

Below is a grammar for *SIMPLE*. Lexical tokens are written in capital letters (e.g., LETTER, NAME). Keywords are between apostrophes (e.g., 'procedure'). Non-terminals are in small letters.

*Meta symbols:*

a\* - repetition 0 or more times of a  
a+ - repetition 1 or more times of a  
a | b - a or b  
brackets ( and ) are used for grouping

*Lexical tokens:*

LETTER : A-Z | a-z -- capital or small letter

DIGIT : 0-9

NAME : LETTER (LETTER | DIGIT)\* -- procedure names and variables are strings of letters and digits, starting with a letter

INTEGER : DIGIT+ -- constants are sequences of digits

*Grammar rules:*

program : procedure +

procedure : 'procedure' proc\_name '{' stmtLst '}'



```

stmtLst : (stmt )+
stmt : call | while | if | assign
call : 'call' proc_name ';'
while : 'while' var_name '{' stmtLst '}'
if : 'if' var_name 'then' '{' stmtLst '}' 'else' '{' stmtLst '}'
assign : var_name '=' expr ';'
expr : expr '+' term | expr '-' term | term
term : term '*' factor | factor
factor : var_name | const_value | '(' expr ')'
var_name : NAME
proc_name : NAME
const_value : INTEGER

```

### Other rules:

1. It is an error to have two procedures with the same name, as well as a call to a non-existing procedure. Recursive calls are not allowed, either.
2. Spaces (including multiple spaces or no spaces) can be used freely, as long as it does not lead to ambiguity, and your tokenizer should deal with that. For example, tokenizer should recognize three tokens 'x', '+' and 'y' in any of the following three character streams:
 
$$x+y, \quad x + y, \quad x \ +y$$
3. SIMPLE programs are always in one-statement-per-line format as in examples.
4. You can make your own assumptions about any other details that have not been specified above – such as case-sensitivity of identifiers/keywords, whether a procedure name may be the same as a variable name or not, etc. Make sure that these assumptions you make are explicitly discussed in your documentation.

## 6 Program design abstractions for *SIMPLE*

The SPA front-end will parse source programs, derive program design abstractions from sources and store them in the PKB. *PQL* query subsystem will then consult the PKB to validate and evaluate program queries. In this section, we address two important issues:

1. How do we specify program design abstractions to be stored in the PKB?
2. How do we refer to program design abstractions in *PQL* queries?

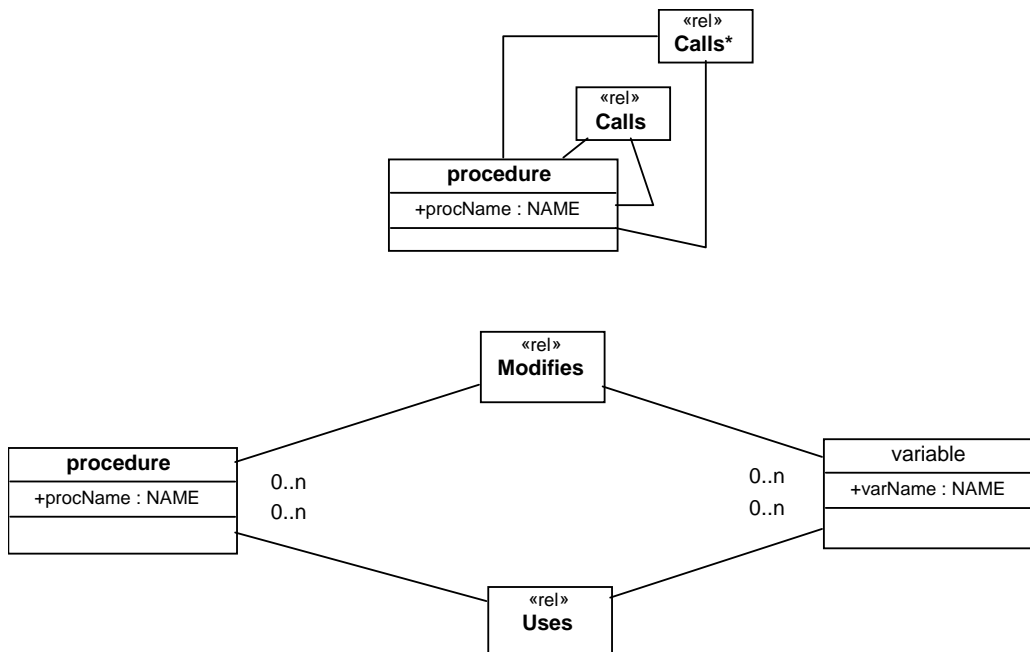
The models described in this section provide the answer to the above questions. Why use models rather than just describe the physical contents of the PKB? Data structures in the PKB may be complex. Writing queries directly in terms of data structures would be unnecessarily complicated while, as we already noted, *PQL* should be easy to use for programmers.

Program design abstractions described in this section are “conceptual” in the sense that they make no assumptions about their physical data representation in the PKB. At the same time, models are sufficient to formulate program queries in a precise but simple and natural way. This is the most important reason for modeling program design abstractions. There are more reasons that make conceptual modeling worthwhile, but as these reasons have to do with the design of a query processing subsystem, we shall defer their discussion until the later sections.

For convenience, we divide program design abstraction models into three parts, namely a *global program design model*, a *syntax structure model* and a *control/data flow model*. The three models are defined in terms of program design entities (such as *procedure*, *variable* or *assignment*), entity relationships (such as *Calls* or *Follows*) and entity attributes (such as *procedure.procName* or *variable.varName*). For each model, you will find a graphical definition (as an UML class diagram) and equivalent textual definition.

### 6.1 A model of global program design abstractions

In the figure below, we see global program design abstractions modeled as UML class diagram.



**Figure 3. A model of global program design abstractions for *SIMPLE* programs**

Program design entities are represented as classes. Relationships among program design entities are represented by class stereotypes <<rel>>. The meaning of relationships is explained below. For any program P, procedures p and q, and variable v:

**Modifies** (p, v) holds if it is possible that procedure p modifies the value of variable v. That is, if variable v appears on the left hand side in some assignment statement in procedure p or v is modified in some other procedure called (directly or indirectly) from p.

The above definition implies that for procedure p below both **Modifies** (p, x) and **Modifies** (p, y) hold:

```

procedure p {
  if x then {
    x = 10; }
  else {
    y = 20; } }
  
```

**Uses** (p, v) holds if it is possible that procedure p uses the value of variable v. That is, if variable v appears on the right hand side of an assignment statement in procedure p or in the condition in p or v is used in other procedure called (directly or indirectly) from p.

**Calls** (p, q) holds if procedure p directly calls q.

**Calls\*** (p,q) holds if procedure p calls directly or indirectly q, that is:

```

Calls* (p, q) if:
  Calls (p, q) or
  Calls (p, p1) and Calls* (p1, q) for some procedure p1.
  
```

Program design entities have the following attributes:

```

procedure.procName (type NAME),
variable.varName (type NAME).
  
```

For example, the following relationships hold in program First of Figure 2 (we refer to procedures and variables via their names given as strings):

**Modifies** ("First", "x"), **Modifies** ("First", "z") -- as there are assignments to variables x and z in procedure First,

**Modifies** ("Second", "i") -- as there is assignment to variable i in procedure Second,

**Modifies** ("First", "i") -- as procedure First calls Second and procedure Second modifies variable i,

**Modifies** ("Third", "v"), **Modifies** ("Second", "v"), **Modifies** ("First", "v"),

**Uses** ("Second", "i"), **Uses** ("Second", "x"), **Uses** ("Second", "y"), **Uses** ("Second", "z"),

Uses ("First", "i"), Uses ("First", "x"), Uses ("First", "y"), Uses ("First", "z"),  
Calls ("First", "Second"), Calls ("Second", "Third"),  
Calls\* ("First", "Second"), Calls\* ("Second", "Third"), Calls\* ("First", "Third").

Your SPA front-end will derive program design abstractions specified by the above model from source programs and store them in the PKB. We are not talking about the data representation for program design abstractions yet.

## 6.2 A model of abstract syntax of *SIMPLE*

Your SPA front-end will build an abstract syntax tree (AST) for a source program. Nodes in the AST are instances of design entities such as assign, expr or while. The following rules model the abstract syntax structure of programs in *SIMPLE*:

Meta symbols: a+ means a list of 1 or more a's; '|' means or

1. program : procedure+
2. procedure : stmtLst
3. stmtLst : stmt+
4. stmt : assign | call | while | if
5. assign : variable expr
6. expr : plus | minus | times | ref
7. plus : expr expr
8. minus : expr expr
9. times : expr expr
10. ref : variable | constant
11. while: variable stmtLst
12. if : variable stmtLst stmtLst

Attributes and attribute value types:

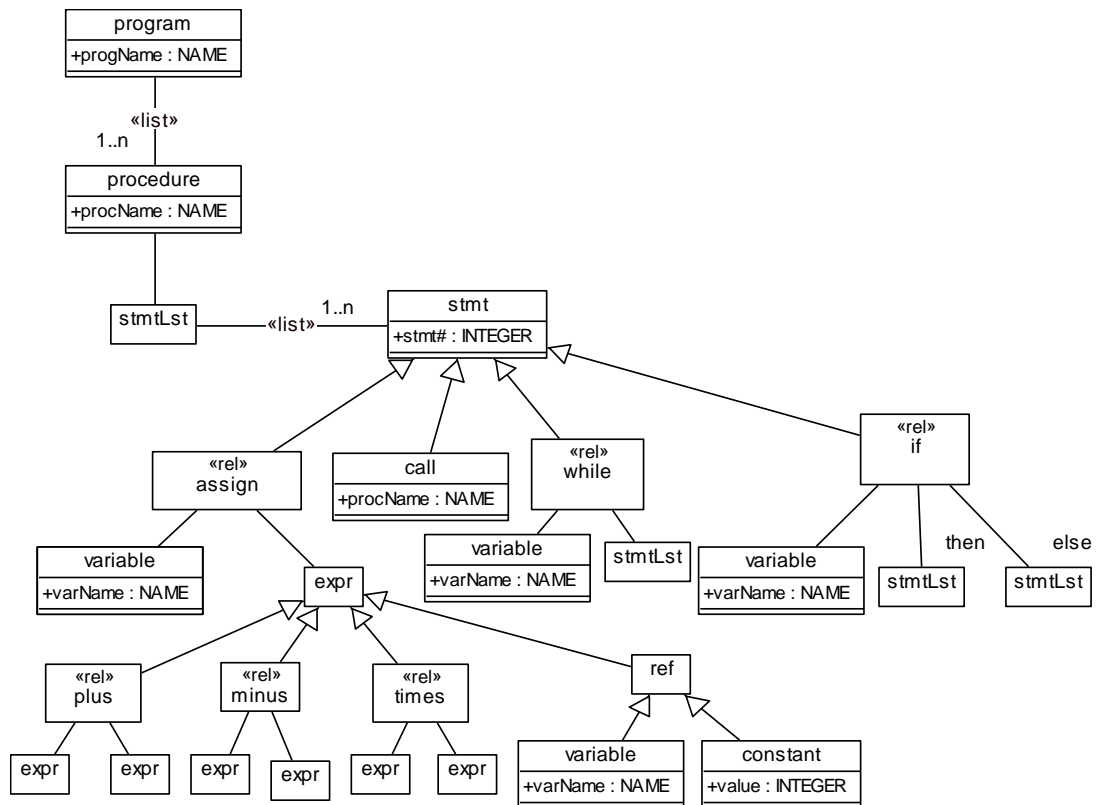
procedure.procName, variable.varName : NAME

constant.value : INTEGER

stmt.stmt# : INTEGER (line number of a given statement, see Figure 2)

**Figure 4. A model of abstract syntax structure for *SIMPLE***

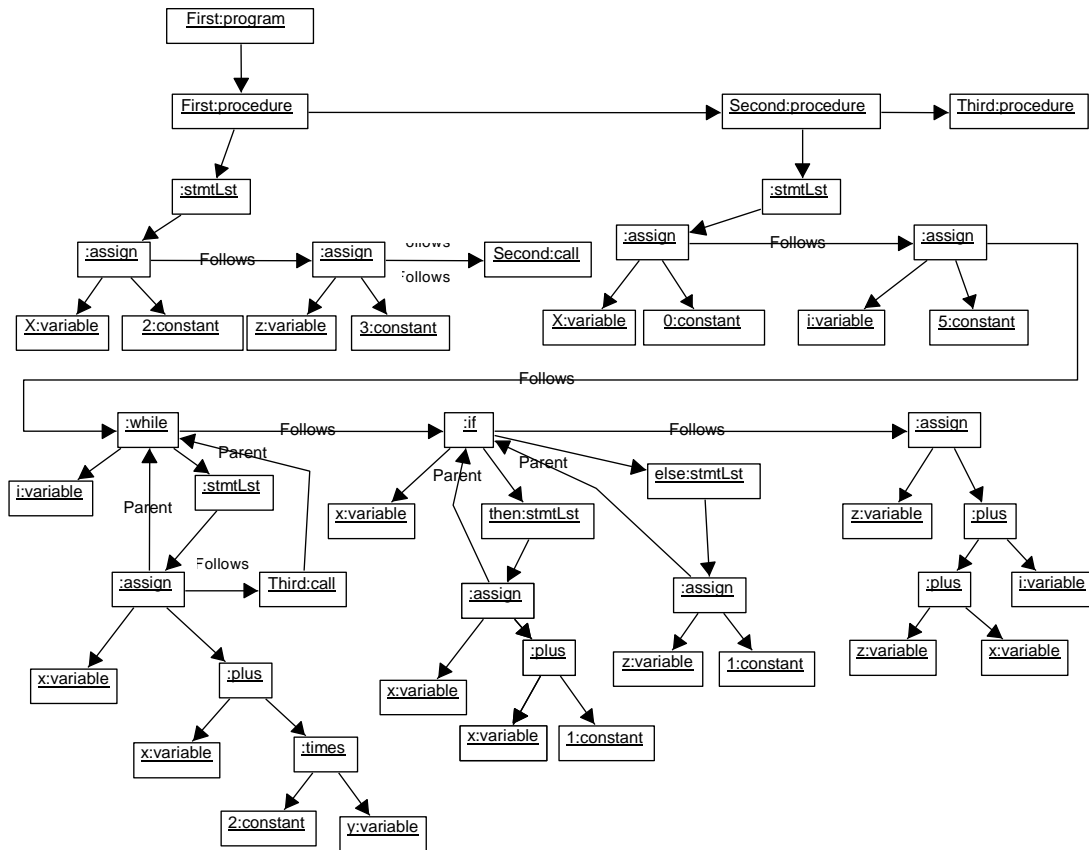
You may like the following graphical model of abstract syntax of *SIMPLE* in UML class diagram notation (Figure 5). Textual definition of abstract syntax (Figure 4) is equivalent to graphical definition (Figure 5).



**Figure 5. Abstract syntax rules for *SIMPLE* as UML class diagrams**

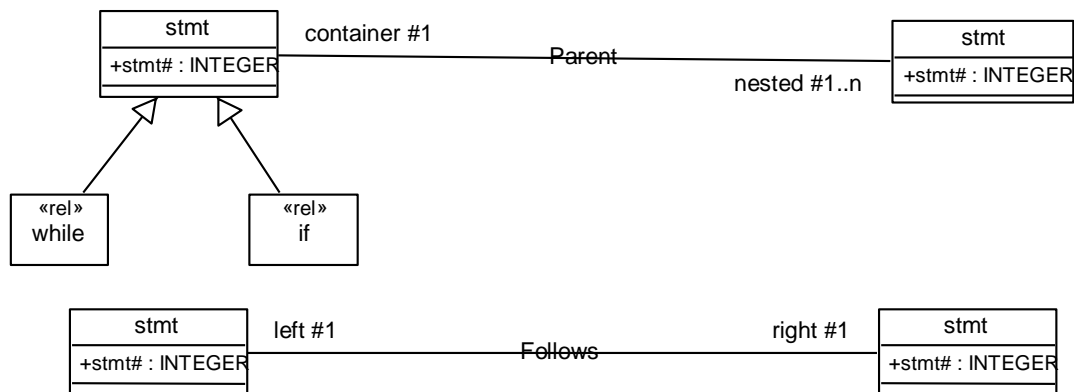
A diagram below shows a partial abstract syntax tree (AST) for program First (Figure 2), built according to the rules defined in abstract syntax models of Figures 4 and 5.

**Note:** Actually, the above is not a standard UML class diagram that you may know of, but it plays its role in explain abstract syntax of *SIMPLE*. Multiplicities are missing on most associations. Associations lack a direction, the name, and the information if it is an aggregation or a composition. The *expr* class exists multiple times, although every class can only appear once in a proper UML diagram. The class *stmtLst* may seem redundant. The main idea with this and other such diagrams or specification artifacts is that you can interpret and understand them. In industry nothing is as perfect as you are used to in the university now. That shall be a main point to take away from this course.



**Figure 6. A partial AST for program First**

We use UML object diagram to depict the AST. Nodes in the AST are instances of design entities in Figure 5. Instance name is reflected at the node using UML convention (e.g., Second:call represents a call to the procedure named Second).



**Figure 7. Relationships Parent and Follows**

Relationships ‘Parent’ and ‘Follows’ describe the nesting structures of program statements within a procedure. Relationships ‘Parent’ and ‘Follows’ are implicitly defined in the AST. For any two statements  $s_1$  and  $s_2$ , the relationship Parent ( $s_1, s_2$ ) holds if  $s_2$  is directly nested in  $s_1$ . Therefore,  $s_1$  must be a ‘container’ statement. In *SIMPLE* there are two containers, namely while and if.

In examples, we refer to statements via their statement numbers, e.g. Parent (3,4). As Parent relationship is defined for program statements, statement number ‘3’ refers to the whole while statement, including lines 3-6, rather than to the program line “while i (“ only. Similarly, statement number ‘7’ refers to the whole if statement including lines 7-9.

For example, in procedure Second (Figure 2) the following relationships hold: Parent(3,4), Parent(3,5), Parent(3,6), Parent(7,8), Parent(7,9), etc.

Relationship 'Parent\*' is the transitive closure of relationship 'Parent', i.e.,

Parent\* (s1, s2) if:

Parent (s1, s2) or

Parent (s1, s) and Parent\* (s, s2) for some statement s.

For any two statements relationship Follows (s1, s2) holds if s2 follows s1 in the statement sequence. 'Follows\*' is the transitive closure of 'Follows', i.e.,

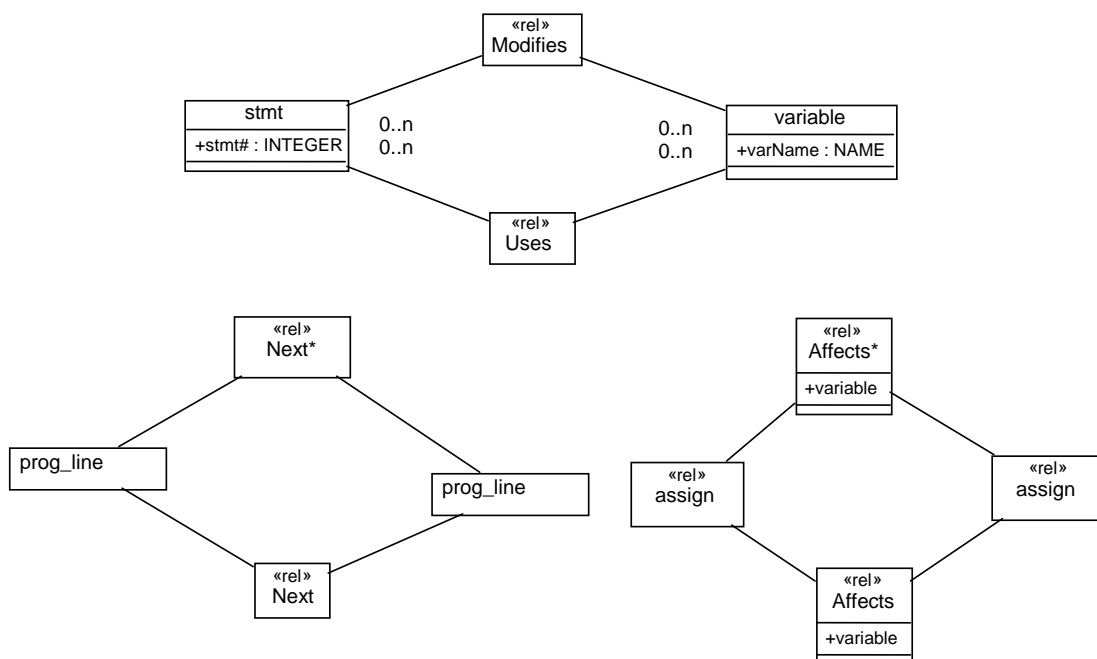
Follows\* (s1, s2) if:

Follows (s1, s2) or

Follows (s1, s) and Follows\* (s, s2) for some statement s.

For example, in procedure Second, the following relationships hold: Follows(1,2), Follows(2,3), Follows(3,7), Follows(4,5), Follows(7,10), Follows\*(1,2), Follows\*(1,3), Follows\*(1,7), Follows\* (1,12), Follows\*(3,12). Statement 6 is not followed by any statement. Notice that statement 9 does not follow statement 8. As in case of the Parent relationship, number '3' refers to the whole while statement and number '7' refers to the whole if statement.

### 6.3 A model of program control and data flow



**Figure 8. A model of program control and data flow.**

For a given statement s, the relationships 'Modifies' and 'Uses' define which variables are modified and used in statement s, respectively. For example, in procedure Second we have:

Modifies(1, "x"), Modifies(2, "i"), Modifies(6, "i")

Uses (4, "x"), Uses (4, "y")

Notice that if a number refers to statement s that is a procedure call, then Modifies(s, v) holds for any variable v modified in the called procedure (or in any procedure called directly or indirectly from that procedure). Likewise for relationship Uses. Also, if a number refers to a container statement s (i.e., while or if statement), then Modifies(s, v) holds for any variable modified by any statement in the container s. Likewise for relationship Uses. For example:

Modifies(3, "x"), Modifies(3, "z"), Modifies(3, "i")

The relationship 'Next' is defined among program lines within the same procedure. Program lines belonging to two different procedures are not related by means of relationship Next.

Let n1 and n2 be program lines. Relationship Next(n1, n2), holds if n2 can be executed immediately after n1 in some program execution sequence.

In case of assignments or procedure calls, a line number also corresponds to a statement. But for container statements, line numbers refer to the "header" of the container rather than the

whole statement as it was the case before. For example, line number ‘3’ refers to “while i {“ and line number ‘7’ refers to “if x then {“.

For example, in procedure Second we have:

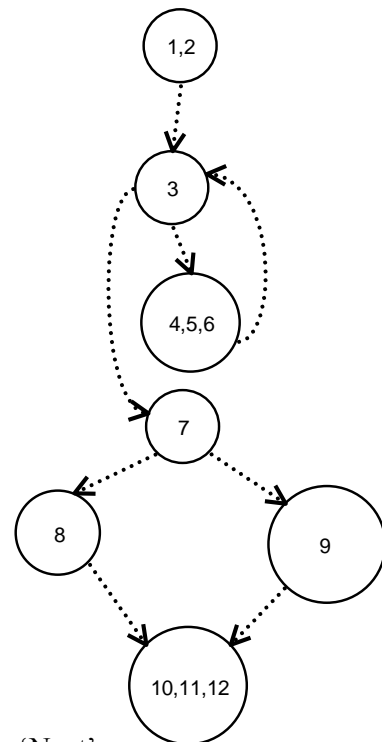
Next(1,2), Next(2,3), Next(3,4), Next(4,5), Next(5,6), Next(6,3), Next(3,7),  
Next (7,8), Next (7,9), Next (8,10), Next (9,10), Next(10,11) and Next(11,12)

Based on the relationship ‘Next’, we can define the control flow graph in a program (CFG). A Control Flow Graph (CFG) is a compact and intuitive model of control flows in a program. Nodes in the CFG are so-called basic blocks. A basic block contains statements that are known to be executed in sequence. That is, a decision point in ‘while’ or ‘if’ always marks the end of a basic block. As statements within a node are known to execute sequentially, we only need to explicitly show control flows among basic blocks.

```

procedure Second {
1. x = 0;
2. i = 5;
3. while i {
4.     x = x + 2 * y;
5.     call Third;
6.     i = i - 1; }
7. if x then {
8.     x = x + 1; }
   else {
9.     z = 1; }
10. z = z + x + i;
11. y = z + 2;
12. x = x * y + z;}

```



The relationship ‘Next\*’ is the transitive closure of relationship ‘Next’:

Next\*(n1, n2) if  
Next (n1, n2) or  
Next(n1, n) and Next\*(n, n2) for some program line n.

In procedure Second we have:

Next\*(1,2), Next\*(1,3), Next\*(4,7), Next\*(3,4), Next\*(4,3), Next\*(6,5), etc.

Relationship ‘Next\*’ describes possible computational paths in a program. It complements relationship ‘Calls’ in the global design model that describes possible sequences of procedure activation.

The relationship ‘Affects’ models data flows in a program. As indicated in the model of Figure 8, relationship ‘Affects’ is defined only among assignment statements and involves a variable (shown as an attribute of relationship ‘Affects’).

Suppose we have two assignment statements a1 and a2 such that a1 modifies value of variable v and a2 uses the value of variable v. The relationship Affects(a1, a2) holds, if the value of v as computed at a1 may be used at a2. That is there must exist at least one computational path from a1 to a2 on which v is not modified.

Notice that relationship ‘Affects’ can only hold between assignments related by means of relationship ‘Next\*’. So as in the case of ‘Next’, “Affects’ may only hold among statements within the same procedure.

In procedure Second we have:

Affects(1,4), Affects(1,8), Affects(1,10), Affects(1,12)

Affects(2,6), Affects(2,10),  
Affects(4,8), Affects(4,10), Affects(4,12)  
Affects(9,10)

But Affects(9,12) does not hold as the value of z is modified by assignment 10 that is on any computational path between assignments 9 and 12.

Suppose you have the following code:

1. x=a;
2. call p;
3. v=x;

If procedure p modifies variable x, that is Modifies (p, "x") holds, then assignment Affects (1, 3) DOES NOT HOLD, as procedure call 'kills' the value of x as assigned in statement 1. If procedure p does not modify variable x; then Affects (1,3) HOLDS.

In the program below, Affects (1,5) and Affects (2,5) do not hold as both x and y are modified in procedure q. Even though modification of variable y in procedure q is only conditional, we take it as if variable y was always modified when procedure q is called. Affects (3,5) holds as variable z is not modified in procedure q.

```
procedure p {  
1.   x = 1;  
2.   y = 2;  
3.   z = y;  
4.   call q;  
5.   z = x + y + z;      }
```

```
procedure q {  
6.   x = 5;  
7.   if z then {  
8.       t = x + 1; }  
   else {  
9.       y = z + x; } }
```

At the same time, we see that the value assigned to z in statement 9 *indirectly* affects the use of z in statement 12. Transitive closure 'Affects\*' caters for this: the relationship Affects\*(a1, a2) holds if assignment statement a1 has either direct or indirect impact on use of v in assignment a2:

Affects\*(a1, a2) if  
Affects(a1, a2) or  
Affects(a1, a) and Affects\*(a, a2) for some assignment statement a.

Consider the following program fragment as an illustration of the basic principle:

1. x=a;
2. v=x;
3. z=v;

modification of x in statement 1 affects variable v in statement 2 and modification of v in statement 2 affects use of variable v in statement 3. So we have: Affects (1,2), Affects (2,3) and Affects\*(1,3).

In procedure Second, we have: Affects\*(1,4), Affects\*(1,10), Affects\*(1,11), Affects\*(1,12), etc.

To compute relationships Affects and Affects\*, you will need to have a CFG and also relationship Modifies for all the procedures.

## 6.4 Summary of program design models

For your reference, here is a summary of program design entities, attributes and relationships for SIMPLE defined in program design models. When writing program queries, we can refer ONLY to entities, attributes and relationships listed below.

*Program design entities:*

program, procedure



stmtLst, stmt, assign, call, while, if  
plus, minus, times  
variable, constant  
prog\_line

*Attributes and attribute value types:*

procedure.procName, variable.varName : NAME  
constant.value : INTEGER  
stmt.stmt# : INTEGER (numbers assigned to statements for the purpose of reference)  
prog\_line.prog\_line# : INTEGER  
call.procName : NAME

*Program design entity relationships:*

Modifies (procedure, variable)  
Modifies (stmt, variable)  
Uses (procedure, variable)  
Uses (stmt, variable)  
Calls (procedure 1, procedure 2)  
Calls\* (procedure 1, procedure 2)  
Parent (stmt 1, stmt 2)  
Parent\* (stmt 1, stmt 2)  
Follows (stmt 1, stmt 2)  
Follows\* (stmt 1, stmt 2)  
Next (prog\_line 1, prog\_line 2)  
Next\* (prog\_line 1, prog\_line 2)  
Affects (assign 1, assign2)  
Affects\* (assign 1, assign2)

## 7 Querying programs with PQL

In this section, we define *PQL* by examples.

### 7.1 General rules for writing program queries in PQL

*PQL* queries are expressed in terms of program design models described in the previous section and summarized in [Summary of program design models](#). So in queries you will see references to design entities (such as procedure, variable, assign, etc.), attributes (such as procedure.procName or variable.varName), entity relationships (such as Calls (procedure, procedure)) and syntactic patterns (such as assign (variable, expr)). Evaluation of a query yields a list of program elements that match a query. Program elements are specific instances of design entities, for example, procedure named “Second”, statement number 35 or variable named “x”.

In a query, after keyword **Select**, you list query results, that is program design entities you are interested to find and **Select** from the program or the keyword **BOOLEAN**. You can further constrain the results by writing (optional) conditions that the results should satisfy. These conditions will include **with**, **such that** and **pattern** clauses. In *PQL*, all the keywords are in bold font.

Here are examples of queries without conditions:

procedure p;

**Select** p

*Meaning:* this query returns as a result all the procedures in a program. You could then display the result, for example, as a list of procedure names along with statement numbers. You could also display procedures along with source code (i.e., the whole program) if you wanted to. The issue of displaying query results is handled separately from the *PQL* by the SPA components *query result projector* and *user interface* (Figure 1).

procedure p;

**Select** p.procName

*Meaning:* returns as a result names of all procedures in a program.

procedure p; assign a; variable v;

**Select** <a.stmt#, p.procName, v.varName>

*Meaning:* query result may be a tuple of program items. Here the result will contain all possible combinations of all assignment statement numbers, procedure names and variable names in a program.

The above query examples did not include any conditions. However, in most situations, we wish to select only specific program elements, for example, you may want to select only those statements that modify certain variable.

You specify properties of program elements to be selected in conditions that follow **Select**. Conditions are expressed in terms of:

- a) entity attribute values and constants (**with** clause),
- b) participation of entities in relationships (**such that** clause),
- c) syntactic patterns (**pattern** clause).

All clauses are optional. Clauses **such that**, **with** and **pattern** may occur many times in the same query.

There is an implicit **and** operator between clauses – that means a query result must satisfy the conditions specified in ALL THE CLAUSES.

Declarations introduce synonyms that refer to design entities. Synonyms can be used in the remaining part of the query to mean a corresponding entity. So you can write:

stmt s;

**Select** s **such that** Modifies (s, "x") -- here we introduce synonym for statement s and constraint the result with condition. As you can guess, in this query, you **Select** all the statements that modify variable "x".

A typical query has the following format:

**Select ... such that ... with ... pattern ...**

In the following sections, we shall introduce queries by examples, referring to program design models explained before.

## 7.2 Types of relationship arguments in program queries

In Section 6, relationship arguments are program design entities such as procedure, statement (stmt), while, etc. Naturally, synonyms of suitable design entities can appear as relationship arguments in queries, for example, we may have Calls (p,q), where p and q are synonyms of program design entity 'procedure'.

In addition, we adopt the following conventions that will allow us to write certain queries in a simpler way:

1. a placeholder '\_' (that is an unconstraint argument) may appear in place of any relationship argument, provided it does not lead to ambiguity. Therefore, Modifies (\_, "x") and Uses (\_, "x") are not allowed, as here it is not clear if the '\_' refers to a statement or procedure.
2. character strings in quotes, e.g., "xyz", can be used for entities that have a NAME, that is, program, procedure, and variable; in such case, a character string refers to an entity with that NAME,
3. integers can be used in place of synonyms 'prog\_line' and 'stmt#'. Such integers are then interpreted depending on the relationship, as a 'prog\_line' in relationship Next, and as a 'stmt#' in other relationships. For example, we can put integer in place of an argument representing a statement (that is a design entity stmt, assign, call, if or while) in relationships Modifies, Uses, Parent, Parent\*, Follows, Follows\*. Similarly, we can put an integer in place of an argument representing an assignment statement in relationships Affects and Affects\*.
4. A synonym 'prog\_line' can appear in relationship that expects 'statement' as argument. In that case, prog\_line is interpreted as a statement number stmt#. For example, the following query is correct:

prog\_line n, n2, n3;

Select n such that Next\*(n, n2) and Parent\*(100, n3) and Modifies(n3, "x") and Affects (n2, 100) and Follows\*(92, n3)

In Parent\*, Modifies and Follows\* n3 is interpreted as stmt#; in Affects n2 is interpreted as a stmt# (it must be assignment).

Your query processor must correctly convert program lines to stmt# in such cases.

On the other hand, we assume that statement numbers do not appear in place of program lines in Next relationship. However, you can allow for that if you wish, and then interpret stmt# as a program line. It will be considered as an extension of the above conventions.

### 7.3 Query examples: querying global program design information

These queries refer to the global program design model described in section [A model of global program design abstractions](#).

Q1. Which procedures call at least one procedure?

procedure p, q;

**Select p such that** Calls (p, q)

*Explanation:* declaration introduces two variables “p” and “q” that can be used in a query to mean “procedure”. Keywords are in bold. This query means exactly: “**Select** all the procedures p **such that there exists** a procedure q that satisfies Calls (p,q)”.

*IMPORTANT:* note that the existential quantifier is always implicit in *PQL* queries. That means, you select any result for which THERE EXISTS a combination of synonym instances satisfying the conditions specified in such that, with and pattern clauses.

*Example:* in program First, the answer is: First, Second

A better way to write this query is:

procedure p;

**Select p such that** Calls (p, \_)

*Explanation:* This query is the same as the one above. Underscore ‘\_’ is a placeholder for an unconstrained design entity (procedure in this case). Symbol ‘\_’ can be only used when the context uniquely implies the type of the argument denoted by ‘\_’. Here, we infer from the program design model that ‘\_’ stands for the design entity “procedure”.

Q2. Which procedures are called by at least one other procedure?

procedure q;

**Select q such that** Calls (\_, q)

Q3. Find all pairs of procedures p and q such that p calls q.

**Select** <p,q > **such that** Calls (p, q)

*Example:* in program First, the answer is: <First, Second>, <Second, Third>

Q4. Find procedure named “Second”

procedure p;

**Select p with** p.procName=“Second”

*Explanation:* Dot ‘.’ notation means a reference to the attribute value of an entity (procedure name in this case).

Q5. Which procedures are called from “Second”?

procedure p, q;

**Select q such that** Calls (p, q) **with** p.procName=“Second”

You can also write the above query in short form, referring to procedure p in Calls via its name:

**Select q such that** Calls (“Second”, q)

Here, we infer from the program design model that the first argument of relationship Calls is design entity “procedure”. By convention, “Second” refers to procedure name. We shall use short form whenever it does not lead to misunderstanding.

Q6. Find procedures that call “Second” and modify variable named “x”

procedure p;

**Select p such that** Calls (p, “Second”) **and** Modifies (p, “x”)

*Explanation:* **and** operator can be used in relationship conditions under **such that** clause and in equations under **with** clause. We infer from the program design model that the second argument of relationship Calls is design entity “procedure” and the second argument of

relationship Modifies is design entity “variable”. By convention, “Second” refers to procedure name and “x” refers to the variable name.

## 7.4 Query examples: querying control and data flow information

Using the same rules, we can write queries about control and data flow relations.

Q7. Is there a control path from program line 20 to program line 620?

**Select** BOOLEAN **such that** Next\* (20, 620)

*Explanation:* BOOLEAN after **Select** means that the result is TRUE (if there exist values for which the conditions are satisfied) or FALSE - otherwise.

Q8. Find all the program lines that can be executed between program line 20 and program line 620.

prog\_line n;

**Select** n **such that** Next\* (20, n) **and** Next\* (n, 620)

Q9. Is there a control path in the CFG from program line 20 to program line 620 that passes through program line 40?

**Select** BOOLEAN **such that** Next\* (20, 40) **and** Next\* (40, 620)

Q10. Which assignment statements are directly affected by variable “x” assigned value at program line 20?

assign a;

**Select** a **such that** Affects (20, a)

*Explanation:* We can refer to program statements via their respective program line numbers.

Q11. Which assignments directly affect a value assigned to a variable in the assignment statement at program line 120?

assign a;

**Select** a **such that** Affects (a, 120)

Q12. Which statements contain a statement (at stmt#='n') that can be executed after line 16?

prog\_line n; stmt s;

**Select** s **such that** Next\*(16, n) **and** Parent\*(s, n)

Q13. Which assignments affect assignment (at stmt#='n') that can be executed after line 16?

prog\_line n; assign a;

**Select** a **such that** Affects\*(a, n) **and** Next\*(13, n)

Q14. Find all statements whose statement number is equal to some constant.

stmt s; constant c;

**Select** s **with** s.stmt# = c.value

Q15. Find procedures whose name is the same as the name of some variable.

stmt s; procedure p;

**Select** p **with** p.procName = v.varName

*Remark:* Notice that the query below is incorrect as ‘prog\_line’ n is not allowed as attribute reference in **with** clause:

prog\_line n; stmt s;

**Select** s.stmt# **such that** Follows\* (s, n) **with** n=10

Similarly, the following query is incorrect either:

**Select** s.stmt# **such that** Follows\* (s, s1) **with** s1.stmt#=n **and** n=10

## 7.5 Query examples: finding syntactic code patterns

Here, we can specify patterns within a procedure code that we wish to find. Code patterns are, therefore, based on abstract syntax model and may involve the following design entities: stmtLst, stmt, assign, call, while, if, ref, constant, variable, expr, plus, minus and times.

Q16. Find all while statements

while w;

**Select** w

Q17. Find all while statements directly nested in some if statement

while w;  
 if if;  
**Select w such that** Parent (if, w)

Q18. Find three while loops nested one in another.

while w1, w2, w3;  
**Select** <w1, w2, w3> **such that** Parent\* (w1, w2) **and** Parent\* (w2, w3)

*Explanation:* notice that the query returns all the instances of three nested while loops in a program, not just the first one that is found.

Q19. Find all assignments with variable “x” at the left-hand side located in some while loop, and that can be reached (in terms of control flow) from program line 60

assign a; while w;  
**Select a such that** Parent\* (w, a) **and** Next\* (60, a) **pattern** a(“x”, \_)

*Explanation:* in a(“x”, \_), we refer to variable in the left hand side of the assignment via its name and in Next\*(60, a) we refer to statement whose number is 60. Patterns are specified using relational notation so they look the same as conditions in **such that** clause. Think about a node in the AST as a relationship among its children. So assignment is written as assign (variable, expr) and while loop is written as while(variable, \_). Patterns are specified in **pattern** clause. Conditions that follow pattern specification can further constrain patterns to be matched.

*Remark:* The two queries below yield the same result for *SIMPLE* programs. Notice also that this might not be the case in other languages. Why?

assign a;  
**Select a pattern** a (“x”, \_)  
**Select a such that** Modifies (a, “x”)

Q20. Find all while statements with “x” as a control variable

while w;  
**Select w pattern** w (“x”, \_)

*Explanation:* We use a place holder underscore ‘\_’ as statements in the while loop body (stmtLst) are not constrained in the query (they are irrelevant to the query).

Q21. Find assignment statements where variable x appears on the left hand side.

assign a;  
**Select a pattern** a (“x”, \_)

Q22. Find assignments that contain expression  $x*y+z$  on the right hand side

assign a;  
**Select a pattern** a (\_, “x\*y+z”)

Q23. Find assignments that contain sub-expression  $x*y+z$  on the right hand side.

assign a;  
**Select a pattern** a (\_, \_“x\*y+z”\_)

*Explanation:* underscores on both sides of the  $x*y+z$  indicate that  $x*y+z$  may be part of a larger expression. \_“x\*y+z” means that  $x*y+z$  must appear at the end of the expression and “x\*y+z”\_ means that  $x*y+z$  must appear at the beginning of the expression. Note that “x\*y+z”\_ means that “x\*y+z” may or may not be followed by other symbols (that is \_ may mean an empty sub-expression).

Q24. Find all assignments to variable “x” such that value of “x” is subsequently re-assigned recursively in an assignment statement that is nested inside two loops.

assign a1, a2; while w1, w2;  
**Select a2 pattern** a1 (“x”, \_) **and** a2 (“x”,\_”x”\_) **such that** Affects (a1, a2) **and** Parent\* (w2, a2) **and** Parent\* (w1, w2)

Q25. Find all statements followed by statement at line 10.

stmt s;  
**Select s such that** Follows\* (s, 10)

## 7.6 Rules for patterns

### 7.6.1 Well-formed pattern specifications

1. Sub-expression must be must be a well-formed expression in SIMPLE (see grammar of SIMPLE), except that brackets are not used. For example, “-x+y”, “x\*y+(z)” “x=2; y=3” are not a well-formed sub-expressions.
2. There can be only one sub-expression specification in a given pattern. For example, pattern a ( \_, \_ ”x” \_ ”y” \_ ) is not valid.

Please check *PQL* grammar at the end of the Handbook.

### 7.6.2 Matching patterns

Matching of sub-expressions takes into account priorities and associativity of operators. Operator \* has higher priority than + and -. Operators + and - have equal priority. Operators are left-associative, meaning the operations of the same priority are grouped from the left to the right. Therefore, \_ “x\*y+z” \_ is an sub-expression of expression “x\*y+z-v” but it is not a sub-expression of “x\*y+z\*v”. Matching of sub-expressions is done on the AST that reflects priorities of operators rather than on the program text.

### 7.6.3 Using an underscore

**Select a pattern** a ( \_, \_ “x\*y+z” \_ )

Underscores on both sides of the x\*y+z indicate that x\*y+z may be part of a larger expression. \_ “ x\*y+z” means that x\*y+z must appear at the end of the expression and “x\*y+z” \_ means that x\*y+z must appear at the beginning of the expression. Note that “x\*y+z” \_ means that “x\*y+z” may or may not be followed by other symbols (that is \_ may mean an empty sub-expression).

## 7.7 Comments on query semantics

Examples in this section are meant to further clarify interpretation of program queries in *PQL*. Read carefully and raise any points that are not clear during analysis (Assignment 1).

### 7.7.1 Implicit existential quantifier in query conditions

The existential quantifier is always implicit in *PQL* queries. That means, you select any result for which THERE EXISTS a combination of synonym instances satisfying the conditions specified in such that, with and pattern clauses.

procedure p, q;

**Select p such that** Calls (p, q)

*Answer:* This query means exactly: **Select** all the procedures p **such that there exists** a procedure q that satisfies Calls (p,q). In program First of Figure 2, the result is procedures First and Second.

### 7.7.2 Implicit ‘and’ operator between query clauses

procedure p, q; assign a;

**Select** <p, a> **such that** Calls (p, q) **with** p.procName=”Second” **pattern** a(“x”, \_)

To qualify for the result, a procedure and assignment statement must satisfy all the conditions specified in the query.

Notice that a query may have any number of **such that**, **with** and **pattern** clauses that can be mixed in any order. There is an implicit **and** operator among all those clauses.

### 7.7.3 Use of free variables

A free variable is not constrained by any condition in the query. Underscore denotes a free variable. You can always replace underscore with a synonym. The following two queries are equivalent:

procedure p, q;

**Select p such that** Calls (p, \_)

**Select p such that** Calls (p, q)

Use of underscore must not lead to ambiguities. For example, the following query should be rejected as incorrect as it is not clear if underscore refers to the statement or to the procedure:  
**Select** BOOLEAN **such that** Modifies (\_, “x”)

#### 7.7.4 A note on meaningless queries

The query evaluator must be prepared to evaluate any query that is formally correct, even if some of formally correct queries may have little use and make little sense. You could try to detect some classes of syntactically correct but meaningless queries during query validation and issue a warning message. But it is risky to reject and impossible to detect all such queries. In any case, decision which queries make sense and which ones do not make sense is somewhat arbitrary. Therefore, your query evaluator must be prepared to deal with all the queries that are syntactically correct.

Here are some example of “strange” queries:

procedure p, q; assign a, a1; while w; variable v; prog\_line n1, n2;  
(the above declarations apply to all the queries below)

**Select** a

*Answer:* all the assignment statements in the program

**Select** a **with** a.stmt# = 12

*Answer:* if statement at line 12 happens to be an assignment statement, then this assignment statement is selected; otherwise, the result is nil. Given numbering as in the program First of Figure 2, the result is assignment statement at line 12.

**Select** BOOLEAN **with** a.stmt# = 12

*Answer:* if statement at program line number 12 happens to be an assignment statement, then this the result is TRUE; otherwise, the result is FALSE. Given numbering as in the program First of Figure 2, the result is TRUE.

**Select** a1 **with** a.stmt# = 12

*Answer:* if statement at program line number 12 happens to be an assignment statement, then all the assignment statements in the program are selected; otherwise, the result is nil.

**Select** w **such that** Calls (“Second”, “Third”)

*Answer:* if procedure Second happens to call procedure Third in the program, then all the while statements are selected; otherwise, the result is nil. Given the program First of Figure 2, the result includes all while statements in the program.

**Select** <a, w> **such that** Calls (“Second”, “Third”)

*Answer:* if procedure Second happens to call procedure Third in the program, then a set of pairs including all the combinations of assignment and while statements in the program are selected; otherwise, the result is nil. Given the program First of Figure 2, the result is a set of pairs including all the combinations of assignment and while statements in the program

**Select** <p, q> **such that** Modifies (a, “y”)

*Answer:* This query means exactly: **Select** all the pairs <p, q> **such that there exists** assignment a which modifies “y”. Though, p and q are not related to assignment a – query is correct. If there is an assignment statement modifying “y”, the result is a set of pairs including all the combinations of procedures in a program.

**Select** BOOLEAN **such that** Calls (\_,\_)

*Answer:* If there is a procedure that calls some other procedure in the program, the result is TRUE; otherwise, the result is FALSE.

**Select** a **such that** Calls (\_,\_)

*Answer:* If there is a procedure that calls some other procedure in the program, the result includes all the assignment statements in the program; otherwise, the result is nil.

## 7.8 A general format for PQL queries

Please refer to Appendix A for a complete PQL grammar. The following is a general format of PQL queries:

select-cl : declaration\* **Select** result-cl ( with-cl | suchthat-cl | pattern-cl)\*

Clauses in rectangular brackets are optional. Asterisk “\*” means repetition 0 or more times. Declarations introduce variables representing program design entities. Synonyms can be used in the remaining part of the query to mean a corresponding entity. The *result-cl* clause specifies a program view to be produced (i.e., tuples or a BOOLEAN value). In the *with-cl* clause, we constrain attribute values (e.g., `procedure.procName=“Parse”`). The *suchthat-cl* clause, specifies conditions in terms of relationship participation. The *pattern-cl* describes code patterns to be searched for.

Notice that a *PQL* query can contain any number of **such that**, **with** and **pattern** clauses and there is a default **and** between any two consecutive clauses. We can swap clauses without changing the meaning of the query.

All the clauses may appear more than one time in a query, in any order:  
**Select ... with ... such that ... with ... with ... pattern ... such that ...**

## 8 Required program quality attributes

In addition to implementing SPA functionalities as described in five assignments, your solution must meet additional quality attributes. In this section, we describe the required quality attributes of your design, program and documentation. You must apply methods described in [Compendium of recommended engineering practices for the project](#) in order to meet the required quality attributes. Further guidelines (particularly, regarding assignment and final report format) are given in assignments.

### 8.1 Flexibility of the SPA

You should design SPA for ease of making changes. In particular, it should be easy to modify SPA in the following ways:

1. extend the source language *SIMPLE* with new language constructs,
2. extend the PKB with new program design abstractions,
3. modify data structures in which you store program design abstractions in the PKB,
4. extend *PQL* with new query formats,
5. refine query evaluator to perform optimizations during query evaluation.

To achieve flexibility, you must design SPA for change. The problem starts when a change you need to do has a global impact on the system and you have to revise lots of code in order to implement the change. So you will design SPA so that the impact of changes is localized to small number of SPA components. You can limit the impact of changes by applying information hiding, careful design of interfaces among SPA components and with table-driven approach.

Design for change is in line with software development life cycle for the project. You will develop the SPA incrementally rather than in one big-bang release. Three development iterations are defined in assignments 3-5. Each new iteration usually will involve some changes to the program base you had built in previous iterations. If you do not plan for change, moving through the iterations will cost you lots of extra work. While you cannot anticipate the impact of change 100%, to some extent you can. Sometimes you may need to refactor your design to accommodate required extensions and to make your programs more flexible. This is a normal process involved in iterative development but of course, you want to minimize the amount of re-work.

### 8.2 Reusability

It should be possible to reuse SPA components (after proper adaptations) in other similar systems, in particular, in SPAs for other source languages and in SPAs using different media for PKB (e.g., a relational database).

### 8.3 Reliability of your programs and testing

Your SPA should be well tested. Testing is an integral part of development. You do not defer testing until the end of the project. Each development iteration should result in a **production**



**quality mini-system.** Each iteration should involve unit testing, integration testing and even some system testing. You should do integration testing after having released every significant piece of work. Why do integration testing often? It will save you lots of work. Integration testing shows if the major components in your system are in sync or not. Integration errors often arise in team project due to miscommunication among team members and making wrong assumptions based on imprecise documentation of component interfaces. Tackle these errors as early as possible. If they go unnoticed to later phases of implementation, it will cause you much time and pain to fix them.

Make a habit to incorporate unit testing into your every day programming. Save test cases so that you can reuse them later when your program changes (regression testing). Use assertions to catch errors as close to the source of error as possible.

## 8.4 Scalability

Your SPA should work for input programs of 1000's lines of SIMPLE code. Make sure that your design decisions scale. In particular, be careful when pre-computing information for query evaluation. We require that relationships Next\*, Affects and Affects\* are computed "on demand" when evaluating queries, rather than pre-computed and stored in memory. The latter won't scale. Solutions that do not scale will affect project grade in negative way.

## 8.5 Quality of program documentation and project reports

Documentation you write must be understandable to other team members, not only to you. Make your documentation clear, precise, easy to follow and always keep it up to date with programs. Adopt documentation and programming standards so that all team members use the same conventions.

Proper documentation of architecture, component interfaces, use of assertions, documentation of test plans and test cases - all these are essential elements of the quality documentation. Quality of documentation contributes much to the quality of assignment and final project reports, based on which we shall grade your work.

## 8.6 Performance of query evaluation

Performance is a major problem in implementing descriptive notations for program queries such as *PQL*. Of course, query evaluation time depends predominantly on the size of the source program. However, design decisions considering the PKB and query evaluator also have much impact on efficiency of query evaluation.

The choice of data representation in the PKB affects the performance of query evaluation. If you design your data structures (for AST, CFG, etc.) taking into account how you will need to compute information during query evaluation, your query evaluator will run faster. Let us say you have stored only forward links among your CFG nodes. Evaluation of the query that requires traversing the CFG backwards will be very time consuming. The way you store information about variables modified/used will have much impact on the evaluation time of queries that involve relationship Affects().

Query evaluator can optimize program queries for better performance. Refer to Section 13.8.3 in "Technical tips" for further discussion of query evaluation and possible optimizations.

## --- Part II: Software Methods ---

### 9 An architecture of an SPA - towards the program solution

By the end of the first assignment, you should have completed requirement analysis and understood SPA's functional and quality requirements. The time comes to think about the program solution that meets those requirements. Often, it is good to try out some implementation during requirement analysis. This early prototyping helps developers gain better understanding of requirements, foresee potential problems and get insights into the design of software architecture. Sometimes, a prototype (or certain parts of it) may be refined and included into the future system. For that reason, we asked you to do implementation exercises in assignments 1 and 2.

#### 9.1 What is a software architecture?

Software architecture provides a master plan for subsequent development phases and a blueprint for a software product itself. Software architecture is critical for coordinating project work and for ensuring that the final product meets its functional and quality requirements. Your task in Assignment 2 is to develop a software architecture for the SPA project.

Many books have been written on software architectures and you may come across many different interpretations of the term "software architecture".

In essence, a software architecture should play the following major roles:

- Role 1. to tell us quickly how major functional and quality requirements of the SPA will be met,
- Role 2. to explain, at conceptual level, how the SPA will work,
- Role 3. to describe SPA components and their interfaces,
- Role 4. to provide the basis for division of work among team members, that is, to tell who is doing what and what are the interfaces among team members,
- Role 5. to provide the basis for planning development iterations.

In great simplification, we can say: architectural design is about decomposing a system into components (at various levels of abstraction) and precisely defining component interfaces. "At various levels of abstraction" means that we may start by identifying large granularity, subsystem level components (such as SPA front-end, PKB, query processing and user interface subsystems in Figure 9) and then continue decomposition of subsystems into lower level components as long as this is needed. For example, parser and design extractor are lower level components in the SPA front-end subsystem.

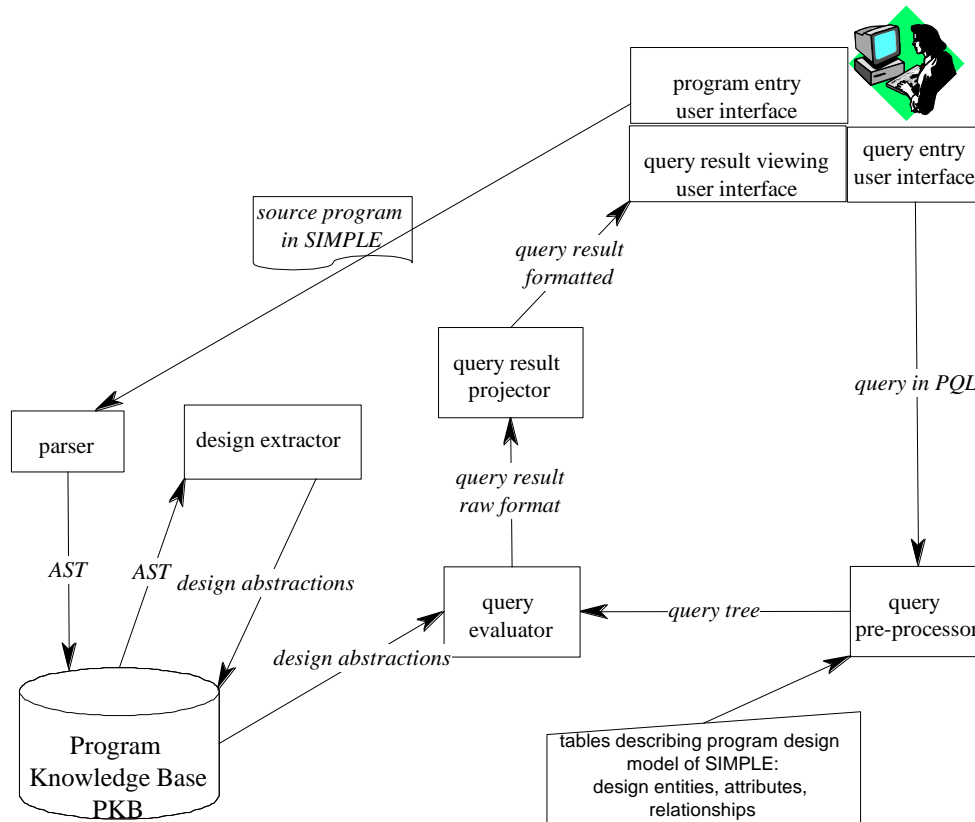
#### 9.2 SPA architecture: component view

The top level decomposition of an SPA is given in Figure 9. We have four subsystems, namely SPA front-end, Program Knowledge Base (PKB), query processing and user interface subsystems. Arrows indicate information flows among components. These arrows provide clues about the nature (not details) of component interfaces. During architectural design (Assignment 2), your main task will be to further decompose the PKB subsystem and provide detailed definition of component interfaces.

Having good decomposition of a system and clear understanding and definition of components interfaces is critical success factor for the project. Interfaces are contracts between components: the interface defines services a component promises to deliver to other components. Having decomposed a system and having defined interfaces, you can describe how your system works (Role 2) and how it meets requirements (Role 1). You can also let different team members work on different components in a fairly independent way. Component interfaces effectively become a communication channel and contractual interface between team members: as long as your team agrees on interfaces, team members can work on different system components. Interfaces ensure that, at the end, components developed by different team members can be successfully integrated. In that way, an architecture with explicit component interfaces makes it possible to divide work among team members (Role 4).

Finally, based on architecture, you can decide what to do first and what to do next during development (Role 5).

1. Your weekly team meetings should focus on discussing component interfaces.
2. Learn how to describe components you are in charge of.
3. Learn to communicate with your team mates in terms of component interfaces rather than in terms of internal implementation details.
4. As the project progresses, refine component interfaces. Keep them always up to date with evolving program.



**Figure 9. Top level decomposition of the Static Program Analyzer**

### 9.3 An SPA front-end subsystem

An SPA front-end builds a PKB. It parses a source program into an AST representation first. In addition to AST, the parser may also produce procedure and variable symbol tables. Then, the design extractor traverses an AST in order to derive other program design abstractions (such as CFG, Calls, Modifies, Uses and other information). Notice that an AST and symbol tables contain complete information about a source program. Once an AST is created, there is no need to analyze program text anymore. Other program design abstractions can be more conveniently derived from the AST rather than by repeatedly parsing the program text.

### 9.4 A PKB subsystem

PKB provides a storage (data structures) for program design abstractions such as AST, CFG, Calls, Modifies, Uses and others. Design abstractions such as AST, CFG, Modifies, Uses are stored in the data structures of the PKB. Design abstractions such as Follows\*, Parent\*, Next\*, Affects and Affects\* are not stored in an explicit form, but rather computed on demand during query evaluation.

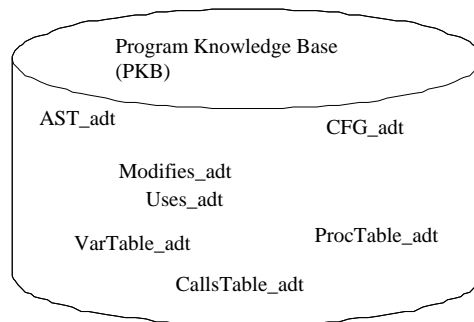
The PKB will expose the API (application program interface) to its clients (e.g., the SPA front-end and the query processing subsystem), while hiding all the data structures used for storing program design abstractions in the PKB. This important design decision is dictated by two considerations:

1. *flexibility and reusability requirement*: the actual data structures to store program design abstractions are likely to change throughout development. Should this happen, you want to minimize the impact of change. PKB API will ensure that changes to data structures in the PKB will not affect the rest of the system.
2. *division of work*: having agreed on the PKB API, work on SPA-front-end, PKB and query processing subsystems can be done in parallel by different team members.

The PKB API plays a central role in the SPA architecture and in your project.

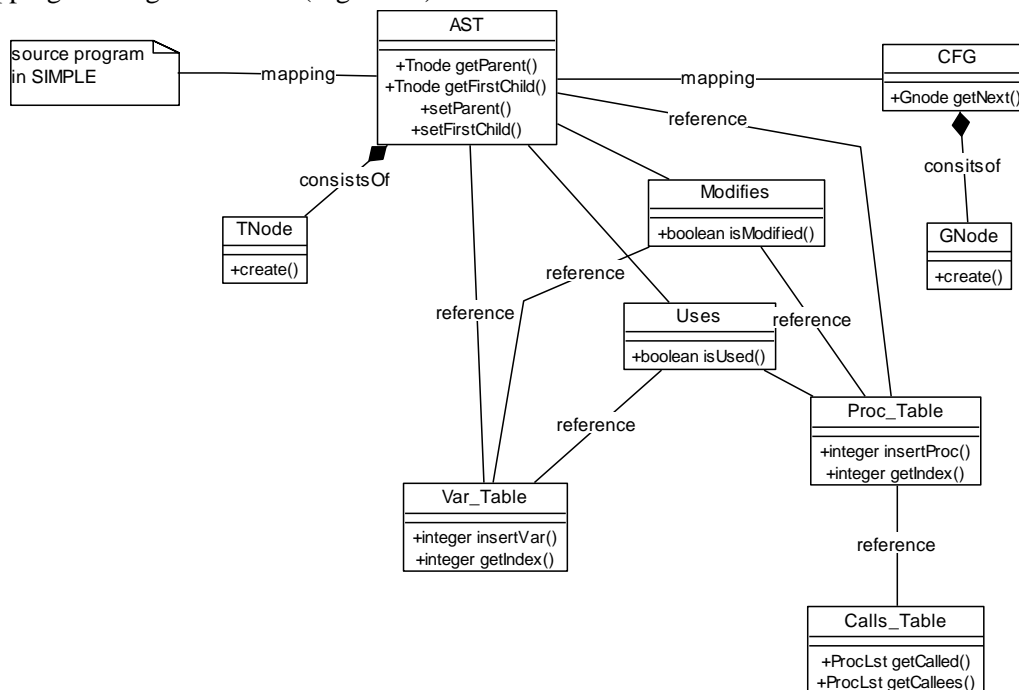
Each of the design abstractions, such as AST, CFG, Calls, Modifies, Uses, etc, is a PKB component, designed as an abstract data type (ADT), using information hiding principle (Figure 10). Therefore, each of those design abstractions will have its own API. The PKB API will be a union of interfaces to its component ADTs. Define interfaces to ADTs one by one and you will come up with the PKB API.

PKB API will also include suitable interface operations to work with design abstractions “computed on demand” such as design abstraction Follows\*, Parent\*, Next\*, Affects and Affects\*.



**Figure 10. Design abstractions as ADTs in the PKB**

It will help you to understand the PKB organization and ADT interfaces if you analyze mappings among ADTs first (Figure 11).



**Figure 11. Associations among ADTs**

The meaning of associations can be described in a table:

Association	The meaning and role of association
mapping among source program and AST	for each statement number we shall be able to identify AST node that contains that statement and vice versa
mapping among AST and CFG	for each AST node we shall be able to identify a corresponding CFG node and vice versa
mappings from Proc_Table to Modifies and from Modifies to Var_Table	for each procedure we shall be able to identify to Modifies vector that indicates variables (mapping to Var_Table) modified in this procedure
etc.	

**Table 1. Documenting mappings among ADTs in the PKB**

## 9.5 Abstract versus Concrete PKB API

PKB API describe how to work with design abstractions stored in the PKB. SPA components will communicate with PKB via PKB API. For example, PKB API will include operations for the parser to create AST nodes, link them to form a tree structure, and operations for Query Evaluator to traverse AST.

You will identify and specify PKB API at two levels: first architecture-level, abstract PKB API (Assignment 2), and then program-level, concrete PKB API (Assignment 3), as you start SPA implementation).

*Abstract PKB API* will be described symbolically, using informal notation, while *program-level, concrete PKB API* will be written in C++, and will consist of public interfaces of classes implementing various PKB ADTs in your program. You will design abstract PKB API first (Assignment 2), and then based on them you will write program-level PKB API (during SPA implementation, starting with Assignment 3). You will need to update abstract PKB API and keep in sync with program-level, concrete PKB API throughout the project.

### 9.5.1 Benefits of abstract PKB API

There are many reasons why in bigger projects engineers chose to work at the level of abstract PKB API before writing program-level PKB API in C++:

1. It is always easier to understand essentials of a problem at hand if we are not overwhelmed with all the details. Abstract PKB API allows us to analyze, understand and document the essential behavior of design abstractions stored in PKB without considering how they will be implemented. We ask, for example “how will Design Extractor build a CFG?” and then identify interface operations for CFG API such as createNode(), setNextLink(FromNode, ToNode), etc. We need not know C++ to do that.
2. As abstract PKB APIs are free of implementation details and expressed in informal (or semi-formal) notation, they are much simpler, shorter and easier to understand than program-level, concrete PKB API.
3. You can implement the same abstract PKB API in many different ways, in many programming languages.

### 9.5.2 How to use abstract PKB API in the project

1. *Communication:* In team projects like yours, abstract PKB API specifications form a contract between team members. Team members communicate in terms of abstract PKB API. Based on abstract PKB API, you agree on how design abstractions will be used, clarify the meaning of various interface operations. You will have less miscommunications in team discussion.
2. *Division of work among team members:* Use abstract PKB API in discussion and assigning specific jobs to team members
3. *Guide to implementation:* When it comes to implementation, operations in abstract PKB API will become public methods (called member functions in C++) of PKB classes. You

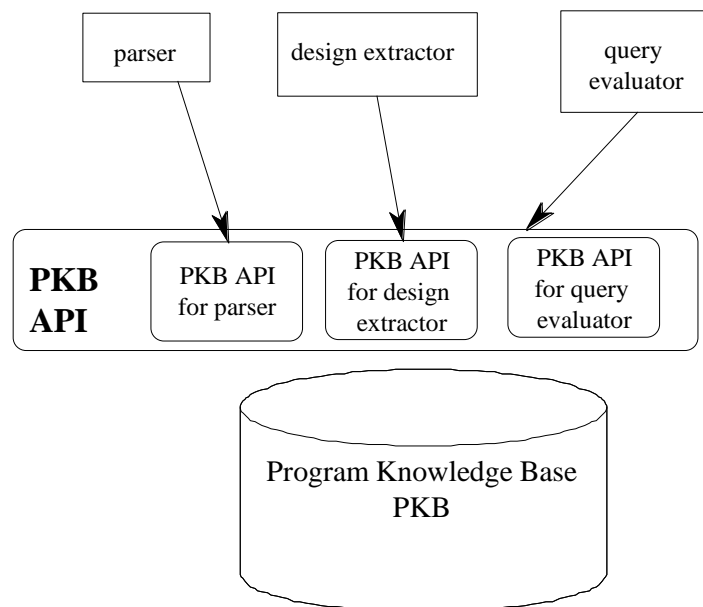
can refine them to add necessary implementation details, but the core concept of API remains the same. You will have less errors in a program.

To observe the above benefits, you need to keep abstract PKB API in sync with implementation. This does not mean that you retrofit implementation details into abstract PKB API. Abstract PKB API will fail to play its role if you did that. But it should be easy to trace from abstract PKB API to corresponding implementation classes (concrete PKB API). Using the same naming conventions in abstract PKB API and corresponding implementation classes (the same names in abstract operations/arguments and member functions), and using the same symbolic names for data types solves this problem. ‘typedef’ can easily assign specific C++ data type to symbolic names. Learn to communicate with your team mates in terms of abstract PKB API rather than in terms of internal implementation details.

## 9.6 How to discover abstract PKB API

PKB API is a union of APIs for all the ADTs in PKB, plus possibly yet other PKB interface operations that do not belong to any ADT. How do we figure out what interface operations a given ADT should have?

To define the PKB API, you will have to understand how different clients (i.e., SPA components) are going to use various ADTs in the PKB. PKB API should make it easy for different clients to create or access information in the PKB without knowing the internals of PKB’s data structures. For the SPA front-end, the PKB API should make it easy to create an AST, CFG and other design abstractions such as Modifies, Uses, etc. For the query processing subsystem, the PKB API should make it easy to traverse the AST and CFG, and retrieve other information stored in the PKB that is required to validate and evaluate queries. You will have to analyze and understand the needs of different clients before you can design a good PKB API (Figure 12).



**Figure 12. Parts of the PKB API for different clients**

The best way to discover APIs for various ADTs in PKB, is to schedule sessions during which team members responsible for different SPA components discuss their needs. A good strategy is to focus sessions on specific program design abstractions such as AST, CFG, Calls, Modifies, Uses, etc. Interface definitions will emerge from those discussions.

The following dialog between a student in charge of Parser and another student in charge of AST illustrates the process of discovering abstract PKB API:

Parser: I need create AST. Can you provide me with simple but flexible means to do that? I definitely do not want to be concerned with how you implement AST.

AST: That's great. I still consider a number of options of how to implement AST. So if you base your decisions on my implementation, you would have to wait. Also, if I decide to change my implementation later on, which is likely to happen we are bound to run into constant problems and lots of re-work. So I will give you AST API - a set of interface operations to AST. You will be able to work with AST using AST API.

Parser: At the moment, you have not implemented AST and I have not implemented a parser. Shall we then just forget about implementation and come up with abstract AST API that logically makes sense, based on common sense understanding of essential properties of AST?

AST: That's fine. We can agree on basic assumptions regarding AST API without being concerned with C++ (which btw I am still learning), doing analysis and initial validation of assumptions "on paper". We'll build a solid base for implementation. I will implement AST API and you will use it to build AST during parsing. So tell me – what you need?

Parser: As I parse a SIMPLE source, sometimes I need create AST node for assignment statement or variable; at other times I want to designate 'variable' node to be a child of 'assign' node; or I want to designate 'variable' node to be a child of 'assign' node or "+" node; or I want to designate '+' node to be the right sibling of 'var' node. The possibilities are endless, I am not in the position to enumerate all those specific cases, it is overwhelming. We definitely need a smart way to do it.

AST: Agree. So how about we start with operation to create AST node, createNode (). It will return to you a reference TNODE that uniquely identifies AST nodes. In that way you will be able to keep track of nodes you created. So I propose operation: TNODE createNode().

Parser: But when I create node I know syntactic entity it represents – whether 'var', 'assign' or constant' – and I would like to record it at the node.

AST: I could give you operations such as createVar(), createAssign(), createConst(), etc. But then we'll have so many operations – it won't be clear or simple. In addition, we'll have to change AST API anytime you change SIMPLE. So how about we add parameter for that purpose: createNode (SYNT\_TYPE syntType), and you provide syntactic type as you create new node?

Parser: Sure, and also we can have operation setRoot (TNODE tNode) to designate 'tNode' as a root of AST.

AST: Ok. Later on when I implement AST API, I will use typedef to assign proper C++ data type to TNODE. We'll have nice traceability between abstract API and implementation, and also any change of the C++ data type will be easy that way.

Parser: How about linking nodes to form a tree?

AST: We'll have a couple of link types: child, sibling, Parent, Follows. For convenience of AST traversal, we may need links up and down the AST, in left and right direction. How about I give you operation in which the first argument will denote the type of the link: CreateLink (LINK\_TYPE link, TNODE fromNode, toNode)?

Parser: That will be fine. We reduce the number of interface operations without compromising convenience or readability. But let's agree and document the exact meaning of each type of link.

AST: Is there anything else that you need?

The dialog continues and at some point AST may want to talk to Design Extractor and Query Evaluator who need to traverse AST and fetch information from AST nodes.

Suppose the responsibilities are as follows:

1. Jane works on the SPA parser,

2. Mary works on the design of AST,
3. Suzan works on the design of CFG,
4. Tom works on the query pre-processor,
5. John works on the query evaluator

You will need to schedule the following meetings:

Team members participating in meeting	The purpose
Jane, Mary	to define interface between parser and AST; what operations are needed to create an AST?
Jane, Suzan	to define the interface between design extractor and CFG; what operations are needed to create a CFG?
Tom, Mary, Suzan	to figure out what information is needed to validate query, to decide if this information is part of the PKB, to define suitable interface
John, Mary, Suzan	to define interface between query evaluator and PKB; what operations are needed to retrieve information from the AST, CFG, etc.

## 9.7 How to document abstract PKB API

It is important to document all the PKB APIs in a standardized way. Here is a suggested format:

<pre> ADT name {   Overview: explain the rationale and responsibility of a component or ADT   Public interface: here you will list interface operations documented as follows   Operation header: returned-value operation-name (list of parameters with names and types)   *Parameters (optional): unless it is not clear from the header, describe parameters   Description: here you describe the effects of the operation (what the operation does) in terms                 of parameters, returned value and whatever else you need to explain the meaning of                 the operation; it is most important that you describe both normal and abnormal                 behavior (handled by assertions and exceptions) }</pre>
--

When you describe abstract PKB API, pay attention to the following recommendations. Refer to the list below often, especially when you work on Assignments 2 and 3:

1. *Overview* clause: Should be brief, but for more complex ADTs, such as AST, it useful to provide some sketches and/or examples. Any new terms or names used throughout operations of a given ADT should be also described in the *Overview*.
2. There is no need to describe obvious parameters under the *Parameters* clause
3. Choice of interface operations for ADTs.
  - a) Interface operations should be well chosen to provide an easy way to work with an ADT.
  - b) The set of interface operations should be complete in the sense that SPA components that work with a given ADT (create it or access information from it) can do so by means of calling ADT interface operations.
4. Adopt detailed standards and use them consistently throughout all PKB API specifications:
  - a) Use symbolic type names (e.g., INDEX, PROC, AST\_NODE, CFG\_NODE, etc. ) rather than specific C++ type names (e.g., int).
  - b) Adopt naming conventions (for operation names, arguments, etc.).
  - c) There should be uniformity in the description of ADTs: similar situations should be described in similar way. For example, VarTable should be similar to ProcTable. ADT Modifies for statements should be similar to ADT Modifies for procedures. ADTs Modifies should be similar to ADTs Uses.
5. *Description* of individual interface operations.
  - a) Keep the set of operations for each ADT simple.
  - b) Interface operation specifications should be readable, understandable to others.



- c) Interface operations for ADTs should be abstract. Use of symbolic type names greatly helps keep interface operations implementation- and programming-language-independent
  - i) *implementation-independent*: should not make unnecessary assumptions about implementation details such as the actual data structures chosen for ADT implementation
  - ii) *programming-language-independent*: avoid specific C++ types
- d) Give names to arguments of interface operations and use these names in the operation *Description*. This helps to describe operations in brief and clear way.
- e) *Description* of interface operations should be simple, concise but precise, unambiguous, complete. In particular, description should specify normal and abnormal behavior of the operation.

An example of VarTable API:

<b>VarTable</b> {
<i>Overview</i> : VarTable is used to keep all the variables that appeared in the program
<i>Public Interface</i> :
<b>INDEX</b> insertVar ( <b>STRING</b> varName); <i>Description</i> : <b>If</b> 'varName' is not in the VarTable, inserts it into the VarTable and returns its index. <b>Otherwise</b> , returns its INDEX and the table remains unchanged. OR: If variable is in the table, returns -1 (special value) and the table remains unchanged.
<b>INTEGER</b> getSize(); <i>Description</i> : Returns the total number of variables in VarTable
<b>STRING</b> getVarName ( <b>INDEX</b> ind); <i>Description</i> : Returns the name of a variable at VarTable [ind] If 'ind' is out of range: Throws: InvalidReferenceException
<b>INDEX</b> getVarIndex ( <b>STRING</b> varName); <i>Description</i> : <b>If</b> 'varName' is in VarTable, returns its index; <b>Otherwise</b> , return -1 (special value)
}

## 9.8 Concrete PKB API (class interfaces in C++ program)

By now you have identified and documented abstract PKB API. Abstract PKB API operations will form the core of public methods (called member functions in C++) in classes implementing design abstractions in the PKB such as AST or CFG. You refine abstract PKB API with all the necessary implementation details. You can also add more interface operations to classes.

Maintain full traceability between abstract PKB API and relevant classes in your program. Follow these two guidelines:

1. Do not retrofit implementation details into abstract PKB API. Abstract PKB API will fail to play its role if you did that.
2. Use the same naming conventions in abstract PKB API and corresponding implementation classes (the same names in abstract operations/arguments and member functions). Use the same symbolic names for data types in abstract PKB API and corresponding implementation classes. 'typedef' can easily assign specific C++ data type to symbolic names.

## 10 Compendium of engineering practices for the project

In this section, we discuss principles and methods that we want you to apply in the project. Many good books have been written on software engineering principles and methods, some of which you studied. We refer you to those books for a comprehensive discussion of software

engineering principles. In this course, we adopt “problem-based learning” approach: we want you to appreciate the value of good engineering principles by applying them to solve SPA problems.

In fact, if you do not follow good engineering approaches, most important of which we discuss in this section, you will not be able to successfully complete the three development iterations and your SPA will not meet functional and quality requirements. You will not score high in this project. We shall insist that you apply methods described in this section so that you can also use them in future projects.

The good news is that you are already familiar with most of the principles and techniques we want you to use. But there is a difference between just being familiar with the concept and knowing how to apply it to your benefit to solve practical problems. In this course, you will learn how to apply good engineering practices to achieve specific project goals.

## **10.1 Preliminaries: common sense software engineering practices**

### *Rule 1: Understand a problem before you go for full-fledged implementation*

Make it a habit to analyze and discuss problem and its solution **BEFORE** you embark on full scale implementation. For this, you need know how to describe essential elements of a problem and its solution in an abstract way (rather than in terms of implementation details). Working with PKB APIs will be great opportunity to learn that (Assignments 2 and 3).

Always focus on the **WHAT** of a problem first (requirements, external view) and think about the **HOW** next (detailed design, implementation). Describe, discuss and evaluate **WHAT** before you decide about **HOW**. Learn to express your thoughts and communicate with other project stakeholders (team mates, project managers, customers, etc.) at the conceptual level. Once you know you have got the concept right, you will come up with better designed, simpler and more stable code.

How do you express the **WHAT** of a problem? You will use two major tools: modeling notations such as UML and component interface specifications. Diagrams and module interfaces convey essential properties of a software system and serve well as a description and communication mechanism during analysis of concepts, design and implementation.

Does it mean that you should not attempt implementing before you have completely understood the problem? No! Experiments and partial implementations are essential to get the understanding of a problem. For example, prototyping an SPA in a very simplified form (Assignment 2) will help you a lot to understand SPA’s essentials. You must be prepared, however, to change (or even scrap) the results of this preliminary implementation work.

We shall now briefly reiterate (we assume you know it!) on two guiding principles, namely separation of concerns and information hiding.

### *Rule 2: Separation of concerns*

Try to deal with one problem at a time. This is the most important principle that can help you in many different ways. We already discussed separation of **WHAT** from **HOW**. In architectural design, you split a system into components so that you can cope with each component in separation from other components. When you model PKB, you concentrate on essential properties (e.g., associations among ADTs in the PKB), describe them in separation from details that are irrelevant at this stage (such as the choice of data structure for AST). You also describe component interfaces in separation from implementation details. Finally, most often (and as it is also the case in this project) you develop system incrementally rather than in one big-bang release. In each iteration, you choose to concentrate on specific set of SPA functions in separation from other functions that you will address in subsequent iterations.

You can separate concerns in both space and time, that is in the software product and in the software process. You can separate concerns horizontally (through decomposition) and vertically (through levels of abstractions). Separation of concerns is indeed one of the fundamental principles in software engineering and in problem solving in general.

### *Rule 3: The virtues of simplicity and standardization*

Let me comment here on one issue that is often overlooked and leads to extra complexities: There are always many ways of solving a given problem, and sometimes there will be also the “best” solution. At first, it seems that we should always go for such “best” solution. However, if we do so we’ll see many different solution patterns across a program. This will increase program complexity. Uniformity (or standardization) of solutions across a program is very important as it makes a program easier to understand and change by many people who work with the program. Many “best“ but incompatible solutions spreading over a program may appear counter-productive at the end. Of course, it does not mean that you should never innovate – there are plenty of problems that require creative and unique solution for a good cause. It all depends on the problem in hand, the importance and value of such solution. In each case, you need be aware if your unique (may be best) solution compromises uniformity and, if it does, there must be good reasons to apply it.

### *Rule 4: System decomposition with information hiding*

Unlike hardware products, software must be flexible, easy to change. Unfortunately, we cannot design software to be flexible in all possible ways. However, if we know up front how the software is likely to change, we can design software for flexibility with respect to those specific types of changes. Take a minute to review Section 8 “Required program quality attributes” to recall flexibility requirements for the SPA.

To achieve flexibility, you apply information hiding as a guiding principle during system decomposition. The idea is to minimize the impact of change. Design decisions that are subject to change must be hidden behind the interface. For example, an ADT in the PKB does not allow its clients (such as SPA front-end and query processing subsystem) to directly access its data structure. Instead, an ADT hides the internal data structure representation and exposes to clients a public interface. For example, an ADT for AST will expose interface operations to create tree nodes, traverse the AST, etc., while hiding the linked list data representation of an AST.

Identifying system components, properly defining component responsibilities, identifying component interfaces and precisely describing them is the essence of software design. Most experienced software engineers are responsible for this task in industrial projects.

## **10.2 Becoming a great architect: Making right design decisions**

Any important design decisions and assumptions are part of software architecture. An essential quality of software developers, and architects in particular, is the ability to make right design decisions. If you are good at it, you will be respected by colleagues and superiors, an indispensable member of a team, and highly rewarded by your employer. If you are the best – you will be a Chief Architect in your company. It takes the right approach, lots of common sense, and experience to become a great architect.

In this course, we show you first steps towards being a great architect: We tell you about the right approach to making design decisions. It is compulsory that you apply this approach throughout the project, and provide evidence (during consultations and assignment/final reports) that you do so. Read this chapter carefully.

Each design problem can be solved in many ways. You make design decisions to chose a design solution that is right for problem at hand, and in sync with other design decisions. You will make 100s of inter-dependent design decisions in the project, and they all collectively should lead to a successful product (SPA).

You make design decisions at all stages of software development as you progress from system requirements to program solution. We group design decisions into these two categories:

### **1. Architectural design decisions**

Software architecture includes system decomposition into main functional components, communication among components, and description of important data repositories. Any other important design decisions that have global impact on your system structure, behavior or

qualities (such as performance, reliability, maintainability and other qualities that matter in a given project) also come under the umbrella of architectural design decisions.

## **2. Detailed design and implementation decisions**

In detailed design decisions, we are mainly concerned with the choice of data representation and algorithms to best meet system requirements. Detailed design decisions are constrained by architectural design decisions. Detailed design decisions can be articulated to some extent in general, programming language-independent way. But often it is good to complement such description with relevant implementation choices, expressed in terms of the underlying programming language (such as C++).

Big O notation is helpful in comparing relative complexity of design alternatives.

### *10.2.1 General approach to making design decisions*

The thinking process behind making architectural and detailed design decisions is very much the same. Here is the general approach to making design decisions:

1. Describe the design problem under consideration
2. Identify, write down and prioritize goals to be met by relevant design solutions
3. Consider alternative design solutions to the design problem under consideration
4. Evaluate each alternative design solution:
  - a) Analyze how well the design solution satisfies identified goals; its strengths and weaknesses
  - b) Analyze any other implications of a given design solution
  - c) Analyze trade-offs among design alternatives

In the above, address any important implications of a given design decision. Depending on the context, take into account complexity of the implementation, performance, how well a given design solution blends with other design decisions, etc.

5. Justify your choice of the preferred design solution.

### *10.2.2 Architectural design decisions in SPA*

Figure 9 depicts high-level architecture of SPA: functional components (such as parser, design extractor or query processing subsystem), PKB as a repository of program information to be shared by SPA components, and major control and data flows among SPA components (shown by arrows). The choices of interface operations in PKB API and their documentation are also an integral part of software architecture.

Architectural decisions in SPA refer to PKB API (interfaces). Important design decisions concern the choice of interface operations for various design abstractions (AST, CFG, etc.) and PKB API documentation standards. Goals and desirable qualities to be met by design decisions related PKB API are discussed in detail in Sections 9.5 - 9.8 of the Handbook. Among most important is the completeness, and ease of use of a set of APIs for a given design abstraction. In documentation of individual operations, most important is readability, completeness of interface operation description (describe normal and abnormal behavior), language-independence of operation description, and consistent use of adopted standards.

### *10.2.3 Detailed design decisions in SPA*

In SPA, detailed design decisions include the choice of data representations for trees (AST), graphs (CFG), Modifies/Uses relationships, and other design abstractions. Then, for algorithms, we have parsing, traversing AST or CFG, computing transitive relationships (Calls\*, Next\*, Affects or Affects\*), and many others.

All you design decisions must collectively target at satisfying SPA's functional and quality requirements in best possible, practical way. SPA answers queries about programs written in SIMPLE. Query answers should be correct (in respect to the specifications provided in the Handbook), and SPA should be able to answer queries fast. Memory consumption during query evaluation is also a concern. SPA should scale to big source programs and complex queries.

Trying to satisfy the above competing design goals is a challenge. There is no "best solution" to satisfy them, but there are many possible solutions that will do that in a better or

worse way. Many detailed design decisions you make will collectively determine the quality of your SPA solutions.

The choice of data representations affects complexity of algorithms. Both data representations and algorithms have to do with memory utilization, and overall complexity of the design solution (and the effort to implement it). The actual time to compute algorithms will also depend on the time to fetch required information from PKB. You need take into account the inter-play among those factors to make right design decisions.

### *Examples of design problems in SPA*

Here are examples of important design problems for which you need make and justify design decisions (it is a small, random sample, which is not meant to exhaustive):

Traversing AST: to answer queries involving patterns, Query Evaluator may need traverse AST many times

Traversing CFG: to answer queries with Next\*(n1, n2), Query Evaluator will have to traverse all the computational paths between n1 and n2 down and up. The choice of data representation for CFG will have impact on efficiency of CFG traversal

Computing Affects (or Affects\*): to answer queries with Affect\*(a1, a2), Query Evaluator will have to traverse CFG, checking sets of modified and used variables on the way. Design decisions regarding CFG and relationships Modifies/Uses will impact efficiency of Affects.

Modifies/Uses: Query Evaluator will often check if a variable is modified/used in a given statement, procedure or CFG node. In other cases, Query Evaluator will have to find all the statements (or procedures) that modify/use a give variable. The choice of data representation for Modifies/Uses will have much impact on query evaluation.

### *10.2.4 Estimating algorithmic complexity: Big O notation*

Big O represents relative complexity of algorithms. You learned about Big-O notation in earlier courses. You are required to use Big O in evaluating and justifying design decisions, whenever it is relevant. In particular, use Big O to represent complexity of computing design abstractions such as Follows, Follows\*, Modifies, Calls\*, Next\*, Affects, Affects\* and others.

Big O should be used in evaluating design decisions together with (not instead of) other factors that matter such as memory utilization, the time to fetch required information from the PKB (e.g., the time to check if a given statement modifies a given variable or not), complexity of implementation, etc.

Be sure that you use Big O clearly and correctly. For example, when you say that algorithmic complexity is  $O(n)$ , say what 'n' stands for. Follow simple rules of how to correctly use Big O described in this [simple summary of Big O essentials](#) by William Shields.

### *10.2.5 Documenting architectural design decisions*

For documenting architectural design decisions related to PKB API you are given precise guidelines in the Handbook. Please follow these guidelines.

Use UML sequence diagrams to document communication among SPA functional components and PKB. Top-level sequence diagram should correspond to the SPA diagram of Figure 9. This diagram can be refined into more detailed sequence diagrams showing communication between SPA functional components and specific data abstractions in PKB.

### *10.2.6 Documenting detailed design decisions*

Detailed design decisions are usually inter-dependent in the sense that you make a number of design decisions that together solve a design problem. Also, the same design solutions may be contributing to a number of design problems. Organize documentation of detailed design decisions to simplify understanding of this situation.

Use one the following description scenarios:

#### **Scenario 1: Many design decisions contribute to solving complex design problem**

In such cases, start with the description of the design problem at hand. Analyze inter-related design decisions that collectively solve the problem. For example, you may want to explain design decisions that led to efficient computation of Affects. Here, "efficient computation of

Affects\*” is your design problem at hand. There are a number of design decisions contributing to Affects such as traversal of CFG and checking which variables are modified/used at CFG nodes. You could explain how design decisions collectively contribute to efficient computation of Affects.

Discuss and compare alternative design solutions that you considered (follow approach discussed in Section 10.2.1). Use Big O notation. Justify your design decision.

### **Scenario 2: Design decision contributes to solving many design problems**

In such cases, start with description of the design solution. Examples of such design solutions is data representation for Modifies (or Uses), and creating a bio-directional map between statement numbers and AST nodes. These design decisions contribute to solving many design problems such as efficient computation of Affects, Affects\*, evaluation queries with patterns, etc.

Discuss and compare alternative design solutions that you considered (follow approach discussed in Section 10.2.1). Use Big O notation. Justify your design decision in view of the many design problems that it contributes to. Make sure that you list those design problems.

**Hint:** Pay much attention to the clarity of documentation of your design decisions. A good practice is to give each important design problem and design solution a unique name (id). Use this name consistently throughout description of your design decisions. Use graphical views to depict inter-relationships among design problems and solutions (i.e., to show which design solutions contribute to which problems, etc.) Clear documentation of design decisions will score you higher grade.

## **10.3 Using UML in the project**

We recommend using UML class, sequence and activity diagrams.

You may want to draw diagrams for a number of reasons: Diagrams can enhance understanding of how SPA works, depict the essence of a general or detail design solution, help you plan testing, or in project management (e.g., in task allocation).

You can draw diagrams at different levels of abstraction, from modeling concepts (requirements), to design, to implementation. Use UML diagrams mainly to model concepts and design (see more recommendations at the end of this section).

Handbook provides many examples of how class diagrams can be used to model concepts in SPA domain: design entities, relationships, AST, and design abstractions in PKB. Activity diagrams are quite natural to use. The rest of the discussion we dedicate to sequence diagrams.

### *10.3.1 Using UML sequence diagrams*

A sequence diagram shows how a group of objects collaborate to achieve *important functions* of a given software system. In business systems, these *important functions* are identified by use cases. Therefore, we may have a sequence diagram per use case.

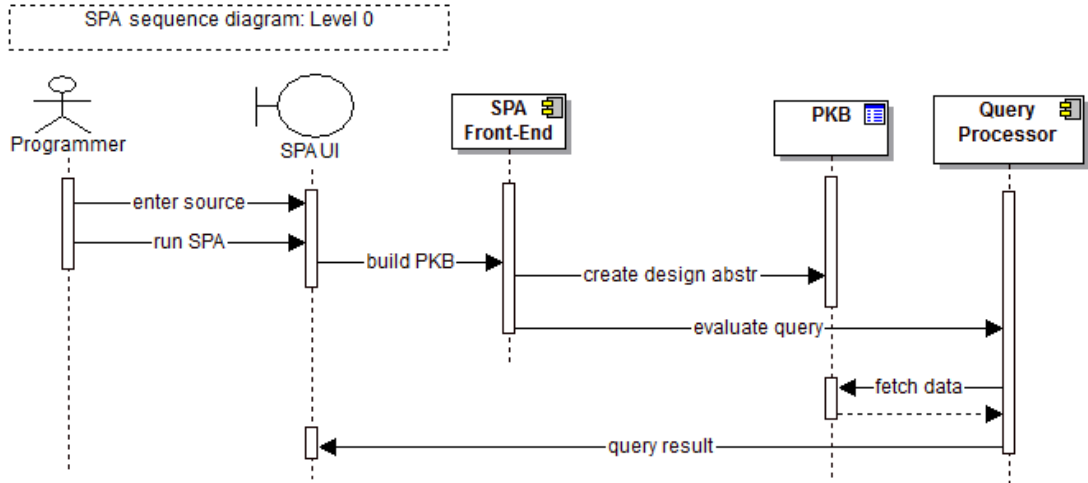
In SPA, queries play the role of use cases. Also, important functions such as computation of Affects may be considered as use case.

Sequence diagrams can be used at all levels of abstraction, from system, to design and to implementation levels. High level sequence diagrams show system-level control flows among major subsystems (e.g., Parser, Design Extractor or Query Evaluator in SPA). Design level sequence diagrams can show control flows (and execution sequences) among components that are assigned more detailed responsibilities (e.g., Query Pre-processor or Query Result Projector).

In SPA, at high abstraction level, we can draw sequence diagrams to show how SPA functional components collaborate to answer queries. Such sequence diagrams have similar contents to high-level component view of SPA architecture shown in Figure 9.

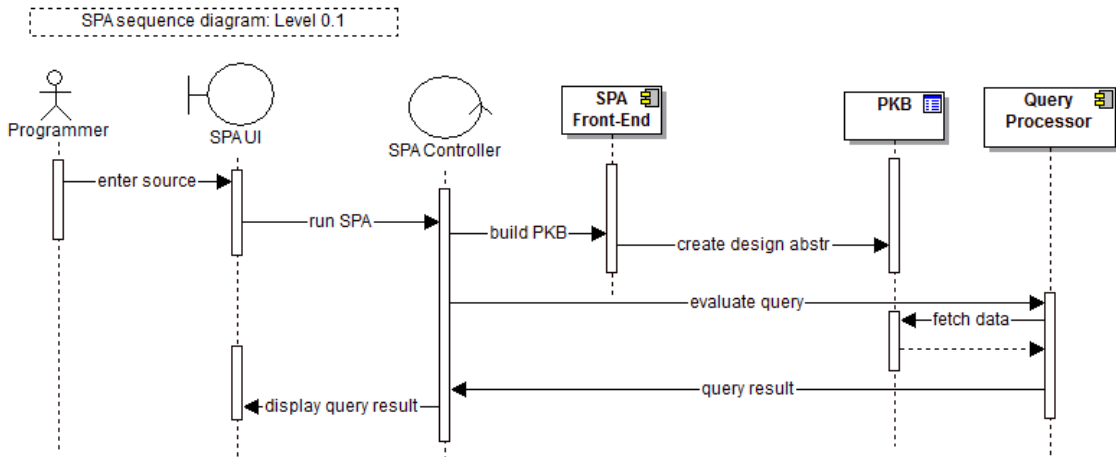
Notice that in sequence diagrams below, different icons are used for user interface elements (SPA UI), functional components (SPA Front-End and Query Processor), and data (PKB).

Labels attached to arrows are responsibilities of relevant components.



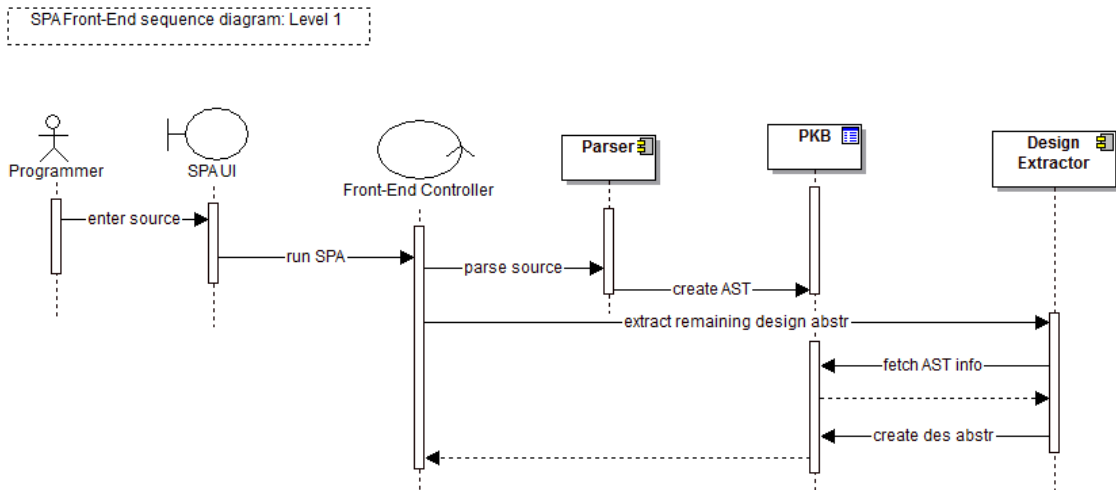
**Figure 13. High-level view of SPA architecture**

Model of Figure 13 is simple and intuitive. Now we can start refining a model of Figure 13 with more details of SPA design. First we observe that a model of Figure 13 lacks proper coordination. Who calls SPA Front-End? Is it right for SPA Front-End to call Query Processor? An important element of good design is to localize control.



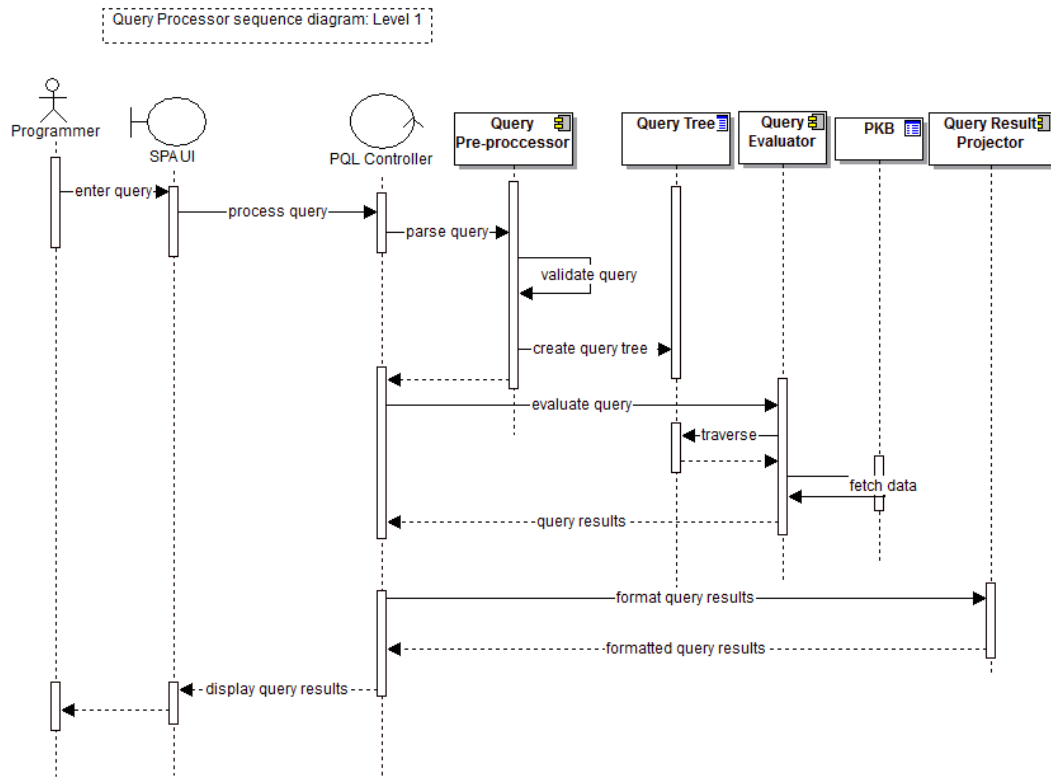
**Figure 14. SPA architecture with Controller**

In model of Figure 14 we introduce SPA Controller to play this role.



**Figure 15. Front-End subsystem sequence diagram**

In model of Figure 15, we decompose SPA Front-End into its two sub-components, namely Parser and Design Extractor (please compare this sequence diagram with SPA model of Figure 9).



**Figure 16. PQL subsystem sequence diagram**

In model of Figure 16, we decompose PQL subsystem.

### 10.3.2 When to use sequence diagrams

- 1) Use sequence diagrams to document your design decisions. They will help you analyze alternative ways to design SPA. All team members will share a clear picture of how SPA works.
- 2) Sequence diagrams will help you understand dependencies among SPA components and tasks assigned to different team members.
- 3) Use sequence diagrams in project planning: Annotate sequence diagram elements with tasks assigned to different team members.
- 4) Sequence diagrams will help you plan and monitor testing. During test planning, annotate sequence diagram elements with information who and when will test different components and their responsibilities. Then, as you progress with testing, mark sequence diagram that have been tested and the degree of their reliability.
- 5) Use models in any way you find useful.

### 10.3.3 Summary of recommendations:

- 1) *Clear purpose:* Draw only diagrams that have a clear purpose. You should be able to explain in what ways a given diagram is useful. For example, diagrams can enhance understanding of the SPA problem, depict the essence of a general or detail design solution, help you plan testing, or in project management (e.g., in task allocation).
- 2) *Use the diagrams in correct way:* Each UML model is meant to model certain aspects of system, but won't be useful if you use it to model other aspects. Use class diagrams to model concepts and structure, but not to model system behavior. Use sequence diagrams to show how objects functional components collaborate to achieve important functions. Use activity diagrams to show algorithmic details of complex computations. Always be sure which diagram to use for a task at hand.



- 3) *Cohesion*: Keep each diagram focused on one specific goal. Do not try to model everything in one diagram. For example, do not model structure (design abstractions in PKB) and behavior (Parser or Query Evaluator) in one class diagram. Not only will you misuse notation (see point 1) above), but your diagram will become unreadable, defeating the very purpose why you spend time to create it.
- 4) *Level of abstraction*: Do not overload diagrams with unnecessary details, as this will blur the picture. Remember that understanding is often the main reason why you create a diagram in the first place. Higher-level diagrams depicting concepts and design are generally more useful than low-level diagrams showing implementation. Implementation-level diagrams are too big and too many, and too difficult to keep them in sync with evolving code.
- 5) *Traceability across diagrams*: Use naming or other conventions to ensure that it is easy to trace model elements across diagrams.
- 6) *Use of UML notation*: Use UML notation and organize diagrams to enhance readability. This is more important than rigorously following textbook examples. In case you do not use diagrams in standard way, always explain what you mean. Use stereotypes if you need extend UML conventions.

## 10.4 Design patterns

Design patterns provide standardized solutions to design problems. You studied some typical design problems for which published design patterns exist in CS2103. If you find instances of those problems in your project, consider applying a suitable design pattern.

By applying a design pattern, you usually win more flexibility, but an overall program solution may be more complex to work with. Always evaluate carefully the trade-offs involved in terms of expected benefits and the cost of applying a design pattern. Apply a design pattern only if the benefits outweigh the cost.

## 10.5 Table-driven technique - query validation example

Table-driven technique is a simple yet powerful way to achieve flexibility. Consider the problem of query validation and the requirement that SPA should be flexible with respect to changes to a program design models. This means that we should be able to easily add new design entities, change attributes of design entities and relationships.

First, let us describe what is involved in query validation. (This function is performed by the query pre-processor in Figure 9). As you enter query (or after you have entered query but before query evaluation), you need to check if the query is valid with respect to program design models for SIMPLE. That means, you need to check if all the references to entities, attributes and relationships in a program query agree with the program design model definition. In particular:

1. all the program design entities in declaration of synonyms should be defined in the program design model. In the case of *SIMPLE*, valid program design entities are procedure, variable, assign, while, etc.
2. all the relationships in query conditions should be syntactically correct. In particular, they should have required number of arguments and the type of arguments should agree to be the same as in the program design model. For example, in relationship Calls (p,q), p and q should be synonyms of the design entity procedure. Cluster (p, s) is not valid as we do not have relationship Cluster () in the program design model. Reference to Calls (p, q, v) is not valid either as Calls expects only two arguments of type procedure.
3. all the references to entity attributes should be valid. First of all, attributes should be defined in the program design model. For example, procedure.value is not a reference to a valid attribute. Equation p.procName=2 in **with** clause is not valid, as the type of attribute procName is string not INTEGER.

How to design query pre-processor so that we can change program design models with minimum effort? You could hard code program design models in the query validation algorithms. The code for query validation will contain switch statements such as:

```
switch ( relationship) {
```

```

    case Calls: expect two arguments; each argument should be either procedure synonym
    or ' _ '
    case Next: expect two arguments; each argument should be either statement number or
    synonym of statement, assign, if, while or ' _ '
    etc.
}

```

In case of changes to the program design model, you will have to find and modify affected parts of the code. Table-driven technique provides a better solution. Rather than hard coding program design models into query validation algorithms, we define program design models in tables. For example:

1	procedure
2	stmtLst
3	Stmt
4	assign
5	etc.

**Table 2. Entity table - EntTable**

Relationship	# arguments	type of arg 1	type of arg 2	type of arg 3
Calls	2	procedure	procedure	nil
Calls*	2	procedure	procedure	nil
Modifies	2	procedure	variable	nil
Modifies	2	stmt, assign, call, while, if	variable	nil
etc.				

**Table 3. Relationship table - RelTable**

An entity attribute table can be defined in a similar way. All the elements of a program design model that you need during query validation are now described in the tables. A query pre-processor will refer to the tables to check whether all the references to entities, attributes and relationships in a query agree with the program design model definition. A query pre-processor will be more generic and much simpler now. Instead of switch statements, you will have the following code:

```

rel = getRelTabIndex (R)
if ( #args ≠ RelTab [rel, 2] ) Error()
if ( arg1 ∉ RelTab [rel, 3] ) then Error()
if ( arg2 ∉ RelTab [rel, 4] ) then Error()
etc.

```

where R is a relationship referenced in a query, #args is the number of arguments of R as it appears in the query, arg1 is the first argument of R, etc. Validation code checks if the actual references in a query agree with their respective definitions in the tables.

The advantage of the table-driven solution is that any changes to the program design models will affect only tables but not the code of the query pre-processor. Changing data in tables is much easier than changing procedural code.

## 10.6 Error handling, exceptions and assertions in C++

Error handling is always a big issue in programming as we all make errors. Exceptions and assertions are means to handle errors in a systematic way. Assertions also play a role in documenting programs.

As there is an overlap between exceptions and assertions, some guidelines can be useful. This section provides such guidelines and pointers to further reading.

Notice that “guidelines” do not mean rules that have to be blindly followed. Sometimes the situation may call for a different solution than recommended by guidelines.

### 10.6.1 Exceptions

“There is an exception to every rule.” Exception mechanism caters for that. Each function has a typical (normal) behavior as well as cases of abnormal behavior. Abnormal behavior signifies that something went wrong and the possibility of an error. You can handle abnormal behavior by writing ‘if’ test and taking appropriate action when the condition is true. But it is better to use exceptions.

It is recommended, that each function defends itself from any possible erroneous calls. In particular, this means that a function should check pre-conditions and define actions to handle situation when pre-conditions are violated (it is called “defensive programming”). The following are advantages of making each function defend itself from possible errors:

1. error handling code is defined in one place – in the function itself
2. callers need not check pre-condition
3. function becomes more reusable
4. program becomes more reliable and easier to change/debug.

Exceptions are written to handle anticipated errors. When an error occurs, an exception code dedicated to the error can either print information that will help you identify the reason of an error in order to fix it, or do some error recovery so that program execution can continue.

**Guideline 1:** Use exceptions to handle cases of abnormal function behavior.

**Guideline 2:** Use exceptions to handle cases of violated pre-conditions.

**Guideline 3:** Throw exceptions as soon as an error is known.

**Guideline 4:** Never assume the correctness of user input. Always check and possibly throw an exception if user input is wrong.

### 10.6.2 Assertions

Assertions are Boolean conditions inserted in code. An assertion states what you believe should be TRUE, according to program specifications, at the program point where you inserted an assertion. As such, assertions are a bridge between program code and requirements. If assertion evaluates to TRUE - the program works as expected; if it evaluates to FALSE - this means that an error occurred. Assertions play an important role in documenting and testing programs.

Usually, program requirements are defined externally to a program. The drawback is that it becomes so easy to change the program without updating the requirements document and vice versa. The power of assertions is that they co-exist with the program code, making it easier to keep specifications up to date with evolving code and vice versa.

Assertions are conditions that express your intention as to what a correct program behavior should be. Macro `assert` is defined in `<cassert.h>`. Whenever assertion condition evaluates to ‘false’ – which means a runtime error – an error message is printed and program aborts. Therefore, assertions cannot be used to recover a program from the error situation.

Assertions are useful for debugging and also as program documentation. Before exceptions were introduced into programming languages, assertions often played the role of exceptions. At which program points should you insert assertions? Any program point that marks a meaningful stage in computation could be annotated with a suitable assertion. Any part of a code (e.g., class method) whose correct execution depends on the program state that can be formulated as a condition could be annotated with an assertion. Insert assertions into your program to test values of critical variables or function parameters to see if they have correct values during execution. In case they do not - print a meaningful error message that will help you localize an error during debugging. Do not remove these error tests from code even if your code works fine, as they will be useful in the future iterations when your program changes.

Checking assertions may affect performance, but assertions are only checked in a debugging mode. You can exclude assertions from a production version by including the line:

```
#define NDEBUG
```

**Guideline 5:** Do not write assertions for cases already catered for in exceptions.

**Guideline 6:** Write assertions to check any conditions that characterize a correct program behavior. Usually, such conditions are implied by the semantics of your program. In particular, post-conditions for functions can be checked by assertions.

For example:

```
// Return parent of input node
TNode* AST::getParent( const TNode* q )
{
// Pre-condition: Make sure input node is not NULL
    assert( q != NULL );

    TNode* p = q->parent;

// Post-condition: Parent can only be NULL, while or if
    assert( ( p == NULL ) || ( p->type == WHILE ) || ( p->type == IF ) );

    return p; }

```

**Guideline 7:** Write assertions whenever you feel they are useful for program documentation.

### **References:**

Here are pointers where you can find exception basics as well as more advanced discussions of using exceptions and assertion in error handling:

[C++ Tutorial](http://cplus.about.com/library/weekly/aa122202a.htm) <http://cplus.about.com/library/weekly/aa122202a.htm>

[C++ Exceptions](http://cis.stvincent.edu/carlsond/swdesign/except/except.html) <http://cis.stvincent.edu/carlsond/swdesign/except/except.html>

[Error Handling with C++ Exceptions, Part 2](#)

## **10.7 Testing your programs**

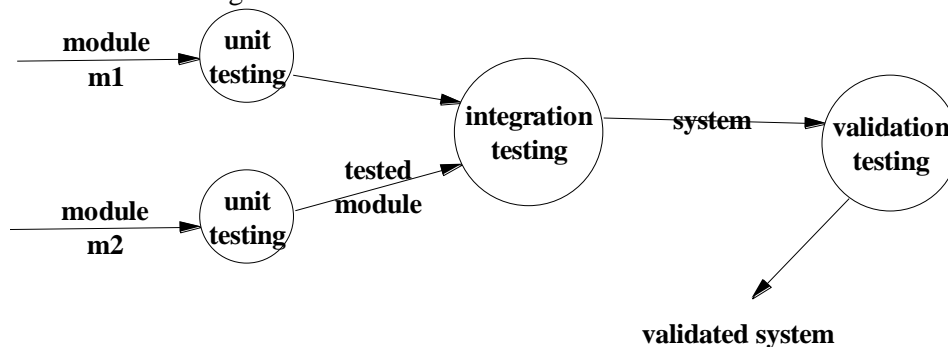
Beyond certain code size and complexity, error-free programs do not exist. You test programs to uncover errors (or to assess the level of program reliability). Reliability requirements for industrial software are higher than for most of the course assignments in university courses. Students often say “we are done!” at first successful compilation and execution of a program for some simple input data. In industry, this is only the beginning and there is a long way before a program is considered finished. That is why in industrial development, testing takes up to 40% of the total project effort.

In this project, we expect program reliability close to industry standards. Allocate enough time for test planning and testing. Make unit testing an integral part of development. Use assertions to document your program and facilitate testing. Allocate time for integration testing and system testing whenever you have wrapped up a piece of work, at least at the end of each of the Assignments 3-5. It is a common practice in successful companies to do integration and system testing on daily basis!

### **10.7.1 You should do the following types of testing:**

1. Unit testing (white-box testing): Testing of individual modules (functions, methods, classes). Try to spot as many errors as possible at the unit level - it will be more time consuming to do this during integration.
  - a) It is responsibility of the developer to do unit testing. Make a habit to incorporate unit testing into your everyday programming.
  - b) You may need to develop test drivers to simulate the environment of the unit under testing.
  - c) Save test cases so that you can reuse them later when your program changes (regression testing).
  - d) Use assertions to insert runtime checks into code. In case of error, print meaningful error message and terminate the program.
  - e) Document the scope of unit testing that you have done.

- f) Make it possible to re-run unit testing after changes.
- 2. Integration testing: testing of a group of components that collectively achieve some higher level function.
  - a) This is primarily validation of component interfaces - does service provider and its clients have the same understanding of the interface operations? You have already tested separately query pre-processor and query evaluator - but do they work fine when you put them together?
  - b) Send random samples of messages to components and validate messages received.
  - c) Perform integration testing a couple of times in each iteration. Integration testing will show if you got the major decisions right. It may indicate possible miscommunication among team members - the earlier the better.
- 3. Validation testing: Checking if the system meets requirements
  - a) Do validation testing at least once at the end of each iteration.



**Figure 17. Testing activities**

### 10.7.2 Test plans and test case documentation.

An effective test case is one that breaks the program - testing is a destructive activity that should be approached with different mind-set than development. Testing should be planned rather than done in ad hoc way. Validation and system testing can be planned as soon as requirement specifications have been completed. If your system passes validation and system testing - your know that you are done.

Test cases must be documented in a standard way and you should be able to repeat tests after changes (i.e., many times during system development and also during future maintenance). Test cases should be stored in a library for future use. Test library must be kept up to date with the evolving program.

Testing is often done by independent groups. So it is essential to document test cases in such a way that others can run them and check the results.

You will be required to produce test plans only for integration and validation testing, documenting each test case in standard was as follows:

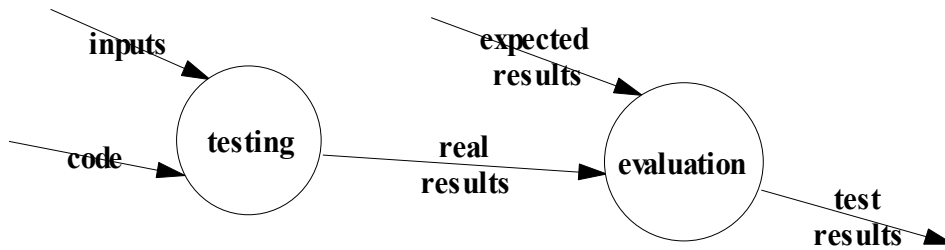
*Test Purpose:* explain what you intend to test in this test case

*Required Test Inputs:* explain what program module (or the whole system) you test and what input must be fed to this test case

*Expected Test Results:* specify the results to be produced when you run this test case

*Any Other Requirements:* describe any other requirements for running this test case; for example, to run a test case for a program module in isolation from the rest of the system, you may need to implement a simulated environment to call the test case.

For unit testing, informally describe the scope of testing and provide samples.



**Figure 18. Running a test case and evaluating the results**

### **10.8 Notes on documenting your programs**

Document your project in clear, easy to follow way. Adopt documentation and programming standards so that all team members use the same conventions. Use UML models to document your design, whenever it useful. Check course web site for recommended tools.

# --- Part III: Software Process ---

## 11 The project team

The design of an SPA involves the following major tasks:

1. building a PKB:
  - a) design of the data structures to store the program design abstractions in the PKB,
  - b) design of the SPA front-end to parse a source program and build an AST,
  - c) design of the algorithms to traverse an AST and derive procedure call, control flow and other program design information, as described in the program design model,
2. designing an application program interface (API) to PKB. PKB API implements all the operations that are needed for its clients, i.e., SPA front-end and query processing subsystem, in particular:
  - a) to traverse ASTs up and down,
  - b) to traverse the CFGs;
  - c) to retrieve program elements related by means of relationships Follow, Follow\*, Parent, Parent\*, Next\*, Affects, Affect\*, as defined in the program design model,
3. designing a query processing subsystem, in particular:
  - a) query pre-processor,
  - b) query evaluator,
4. design of the user interface subsystem.

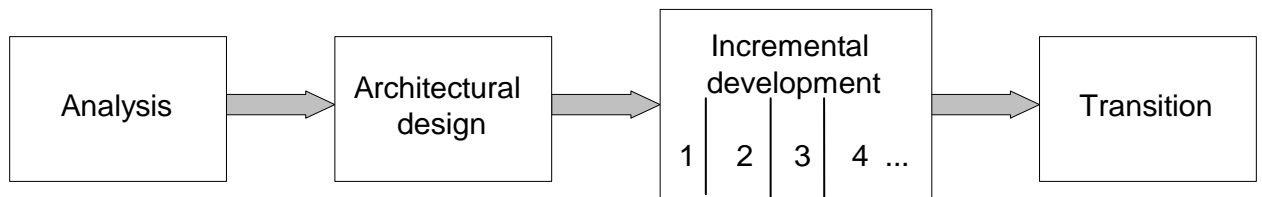
You will form teams of 4-6 students. Each team will be divided into 2 groups (2-3 students per group), called Group-PKB and Group-PQL, respectively. Roughly, Group-PKB will be responsible for [task 1](#), Group-PQL will be responsible for [tasks 3 and 4](#), and both groups will be equally responsible for [task 2](#) - designing a PKB API.

The PKB API is critical to the success of the project. It must be well designed, taking into account data representation in PKB and the needs of the query evaluator. Documentation of PKB API will be the major deliverable from Assignment 2. But you will not get it right at the first shot. So be prepared for refinements as project progresses. Members of the Group-PKB and Group-PQL will have to communicate a lot. To facilitate communication, it is essential that you document PKB API in a clear way and keep the document up to date with changes throughout the project.

## 12 An SDLC for the project

To manage project complexity, you will develop the project in iterations, incrementally. Each new iteration will extend the program implemented in the previous iteration. The main purpose of iterative development is to tackle difficulties one by one. In this section, we discuss motivation and basic rules for incremental software development lifecycle (SDLC).

The SDLC for this project is a variant of the Unified Process. Read brief introduction to the Unified Process in Chapter 2, M. Fowler *UML Distilled, Second Edition*, Addison-Wesley, 2000. The SDLC and development iterations for your project take into account specific technical difficulties involved in the design of an SPA. Project phases include analysis, architectural design and construction (incremental development).



**Figure 19. SDLC phases for the project**

## 12.1 Analysis

During analysis, you get an understanding of:

- what product you are going to build, and
- how you are going to build it.

Analysis is risk-driven - you want to identify risks and address them early in the project. Risks can be related to requirements (e.g., unexpected changes, fuzziness), technical issues (e.g., complexity of algorithms), personnel (e.g., team member gets sick or lacks of expertise in some area) and, sometimes, politics (unlikely in your case). First, you need to analyze the problem description to be sure that you understand all the concepts. If you were implementing a business application, you would have to interview users, perhaps build screen prototypes to clarify requirements. You would create models of required system functions (such as use cases) and conceptual models (such as class diagrams). In this project, the problem has been described for you. But as the problem is complex, you will produce program design abstractions for sample programs by hand and play with them to get an understanding of the concepts. You will also evaluate by hand program queries for sample programs.

## 12.2 Architectural design

During architectural design you will do the following:

1. identify major components of the SPA and describe their responsibilities,
2. describe how components will work together,
3. document component interfaces,
4. model associations among ADTs in the PKB as UML class diagram,
5. describe in detail the meaning of associations.

The most important part of architectural design will be to define interfaces to major program design abstractions (such as ASTs, CFG and procedure call trees). These interfaces should hide physical representation of data and provide a convenient set of operations to retrieve information from the PKB during query evaluation. The union of interface operations for all program design abstractions stored in the PKB will be for the PKB API.

At the end of architectural design, you should be also clear about who will do what in a team and how you will communicate within the team during development.

## 12.3 Incremental development

This is a construction phase in which you will build an SPA incrementally, in three iterations. At the end of each iteration, you will produce a fully tested and integrated program, implementing part of system functionality. Each iteration can be viewed as a mini-project, as it will involve some amount of analysis, design, implementation, integration and testing.

Incremental development is the key to successful development of complex and large systems. It is based on separation of concerns and “divide and conquer”, principles that have been effectively used for years to tackle complex tasks in various engineering disciplines.

Often, a program you develop in one iteration will tackle only a simplified problem. Therefore, you will have to extend the program in the later iterations. Program flexibility and design for change with information hiding are essential in incremental development. This is not a limitation in this project as one of your goals is to design a flexible SPA that can be easily changed in many aspects. Planning for change from the beginning will only help you achieve this goal.

Obviously, the order and scope of iterations must be carefully defined to ensure smooth project development. Therefore, incremental development starts with planning. It is important to plan iterations so that:

1. problems attacked during each iteration are not overly complex, and
2. program solutions developed in earlier iterations can be easily reused and extended in later iterations.

You will be given a detailed plan for development iterations in Assignments 3, 4 and 5.

Having read the description of the SPA, you may have an impression that it must a formidable task to implement it. In fact, in the scope as it was described - you are right, it is a



large and difficult task, especially if you think about making it in one big-bang release. But the degree of difficulty depends on how you scope the problem. Suppose, for example, that you further simplify *SIMPLE* and assume that a program consists of a single procedure with a sequence of assignments such as  $x = 1$  or  $x = y$ . In addition, you will only address queries of the following formats:

stmt s1, s2;

**Select** s1 **such that** Parent (s1, s2) **with** s2.stmt# = int?

**Select** s1 **such that** Follows (s1, s2) **with** s2.stmt# = int?

and answer them for different values of the argument ‘int?’ (i.e., statement numbers). Does it sound difficult? You will find out soon as we shall ask you to develop such an SPA prototype at the beginning of the project. If it looks too trivial to you - you can extend the scope of the prototype by addressing while or if control structures or by addressing queries in the following format:

assign a1, a2; variable v;

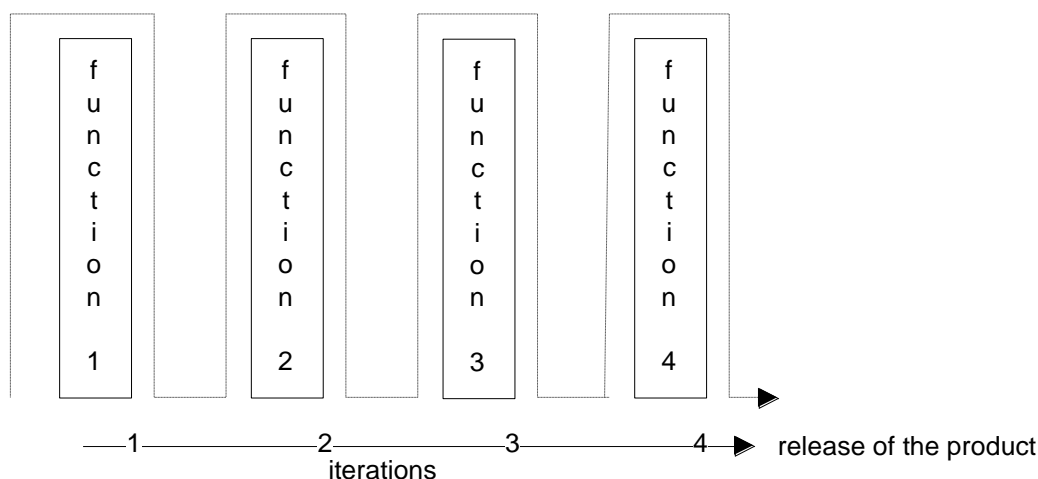
**Select** a1 **such that** Affects\* (a1, a2) **with** a2.stmt# = ? **and** v.varName = ?

## 12.4 Planning incremental development

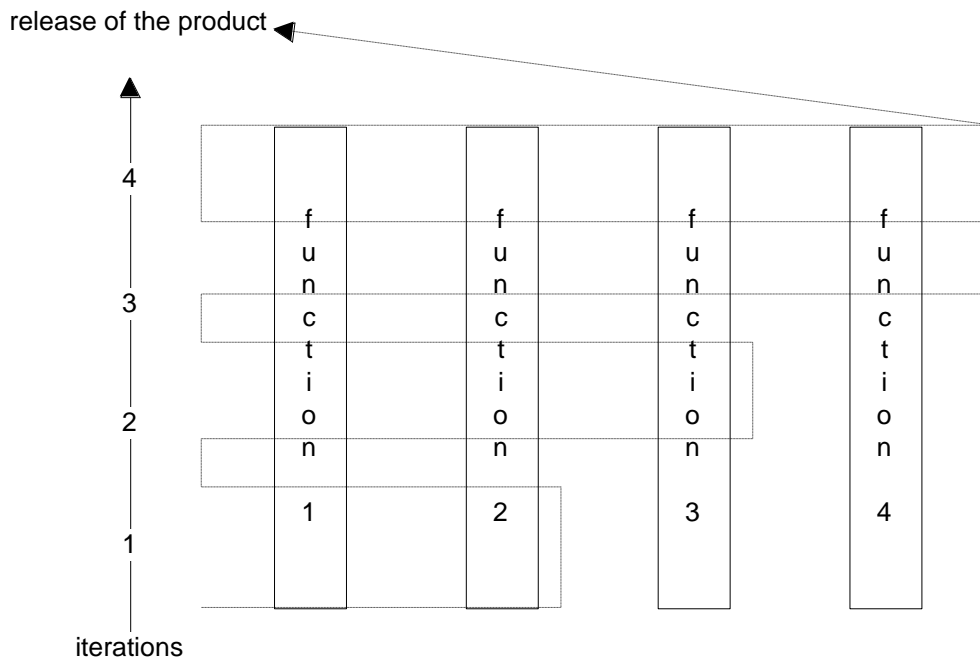
Iterations mark project stages at which programs reach certain level of maturity. While not fully stable, programs released at the end of each iteration are designed to facilitate changes that occur in subsequent iterations. Having analyzed a problem and designed a software architecture (Assignments 1 and 2) - how do you plan development iterations? You plan iterations based on:

- system requirements, i.e., functions to be implemented, and
- major internal structures that are required to implement a solution.

Let us compare the following two iteration strategies. Let us call the first strategy the depth-first strategy. In each iteration, you identify a problem (e.g., a system function you need to implement) and try to provide a complete program solution for that problem in one iteration (i.e., you go deep into the problem). In subsequent iterations, you add more functions one by one and you come up with the complete system at the end. For example, Group-PKB could start by developing a parser for *SIMPLE* and generating an AST for programs. At the same time, Group-PQL might develop a strategy for query validation and start designing a general mechanism for query evaluation.



**Figure 20. Depth-first incremental development**



**Figure 21. Breadth-first incremental development**

Here is another idea for incremental development that we shall call the breadth-first strategy. Rather than digging deep into specific narrow problems, in each iteration we touch on and experiment with a number of related problems, but in a simplified form. We move broad and shallow through the problem and solution spaces rather than deep and narrow as before. Like in the example in the previous section, Group-PKB may start with a very small subset of *SIMPLE* and develop PKB containing a wide variety of program design abstractions for that subset. Group-PQL can start by implementing specific program queries, for the same subset of *SIMPLE* that Group-PKB covers. In subsequent iterations, Group-PKB would extend subset of *SIMPLE* and refine the data representation for program design abstractions in the PKB.

The following steps illustrate possible iterations in each strategy:

*Depth-first iterations:*

1. develop parser for *SIMPLE*
2. generate an AST
3. develop AST traversal algorithms
4. generate CFG
5. parse queries, etc.

*Breadth-first iterations:*

1. select small subset of *SIMPLE*  
     develop parser for subset, simplified AST, CFG  
     validate/evaluate some simple types of queries
2. extend the subset of *SIMPLE*  
     extend AST and CFG, add more program design abstractions to PKB  
     address more types of queries
3. ...

Which strategy is more appropriate for the SPA project? In Assignment 2 we shall ask you to think about the trade-offs and to plan a strategy for incremental development of the SPA project.

## --- Part IV: Technical Tips ---

### 13 Technical tips

#### 13.1 User interface to SPA

User interface consists of three parts, namely source program entry, query entry and query result display. User interface will be very simple.

#### 13.2 Parsing *SIMPLE*

You can use the technique described here to parse *SIMPLE* programs and *PQL* queries. For *PQL* queries, refer to section [Entering program queries \(Group-PQL\)](#) first.

Start with the following subset of *SIMPLE*:

```
program : procedure
procedure : 'procedure' proc_name '{' stmtLst '}'
stmtLst : stmt ';' (stmt ';')*
stmt : assign
assign : var_name '=' var_name | const_value
constant : INTEGER
```

You are free to choose parsing technique (for example, you may choose to implement parser in Perl). Here, we describe an easy way to write a predictive recursive descent parser. For each grammar symbol (such as program, procedure, etc.) have a procedure that attempts to recognize a corresponding program structure. Procedure GetToken () reads and returns a token. As the language rules are simple, we can determine what to do (i.e., which procedure to call) based on the next token (stored in variable next\_token). Assume that source programs in *SIMPLE* are syntactically correct - **Group-PKB does not have to deal with error handling when parsing source programs**. In case of error - just stop parsing. Here is pseudocode:

```
Match (token) {
    if (next_token = token)
        next_token = GetToken ()
    else
        Error ()
}
```

```
Program () {
    next_token = GetToken();
    Procedure () }
```

```
Procedure () {
    Match ('procedure');
    Match (proc_name);
    Match ('{');
    StmtLst ();
    Match ('}');
}
```

```
StmtLst () {
    Stmt ();
    Match (';');
    if ( next_token = '{' )
        stop;
    else
        StmtLst ();
}
```

```
}
```

```
Stmt () {  
    Match (var_name);  
    Match ('=');  
    Match (var_name or const_value);  
}
```

Once you have implemented and tested a parser for the above subset of SIMPLE, it will be easy for you to extend it with other language constructs.

In case of error, your parser should print an error message and stop parsing. However, your parser should allow one to freely use spaces as separators in a program. For example, the parser should accept any of the following statements:

```
x=a+b;  
x = a + b ;  
while i{  
    x=0;}  
while i {  
    x=0 ; }
```

### 13.3 Notes on AST

You will need to inter-link AST nodes for ease of traversing an AST up and down. Each AST node should include at least the following three links:

firstChild link	rightSibling link	up link
-----------------	-------------------	---------

You may find some extra links useful. Think about suitable representation of relationships 'Parent' and 'Follows'. Should you have links for them in the AST or would you rather compute them during query evaluation based on the above three links? What are the trade-offs? Definitely the transitive closures 'Follows\*' and 'Parent\*' should be computed on demand during query evaluation.

### 13.4 Generating an AST during parsing

Before you make your parser generate an AST for the source program, extend parser's pseudocode with AST generation actions. Not only will this enhance your understanding of the AST generation logic before the actual implementation, but it will also help you convey to the PKB designers your needs with respect to parser's interface to the AST. Here is a sample:

```
TNode* StmtLst () {  
    TNode *curNode, *nextNode;  
    curNode = Stmt ();  
    Match (;);  
    if ( nextToken = '}' ) return curNode;  
    else {  
        nextNode = StmtLst ();  
        curNode -> setRightSib (nextNode);  
        return curNode; }  
}  
TNode* Stmt () {  
    Tnode *assign, *leftVar, *expr;  
    assign = new ('assign');  
    Match (var_name);  
    leftVar = new (var_name);  
    assign -> setFirstChild (leftVar);  
    Match ('=');  
    Match (var_name or const_value);  
    expr = new (var_name or const_value);  
    leftVar -> setRightSib (expr);
```

```

    return assign;
}

```

### 13.5 Creating symbol tables during parsing

Your parser should store procedure and variable names in the symbol tables. All the references to procedures and variables in the AST, CFG and in other program design abstractions should be replaced by indices to the symbol tables. A procedure table (ProcTable) should contain procedure name, reference to the root of AST, reference to the CFG entry node, reference to the vector of modified variables, etc.

Index	procedure name	AST root	CFG entry	Modified	Used
1	First				

**Table 4. ProcTable**

Symbol tables will nicely integrate program design abstractions in the PKB. By consulting the ProcTable, you will be able to tell which procedures a program consist of, where are their respective ASTs and CFGs, etc.

Store information about which procedures call which other procedures (i.e., Calls relationship) in the CallsTable. You will not store relationship Calls\*, rather you will compute Calls\* information as needed, based on CallsTable.

### 13.6 What information should be stored in the PKB?

It is not practical to store all the program design abstractions specified above in the PKB. Some of the information will be computed “as needed” from the PKB, when you evaluate queries. Your PKB should contain: an AST, global program design abstractions and CFG (Next relationship). Transitive closures of relationships will be computed “as needed” from PKB during query evaluation (the PKB API should provide suitable interface operations to do so). You will have to decide whether relationship Affects() should be pre-computed and stored in the PKB or rather also computed “as needed”.

What are the factors to consider and trade-offs in deciding if some information should be stored in PKB or computed “as needed”?

### 13.7 Relationships Modifies and Uses

To each procedure, statement and statement block (in while loop body, if-then and if-else), you will need to attach a corresponding set of variables modified and used.

Consider bit vector as a compact and easy to manipulate implementation of the relationships Modifies and Uses. As you will store program variables in the symbol table, each variable may be referred to by a unique number, an index in the symbol table. Represent relationship Modifies (s, v) as a bit vector in which i’th bit is 1 if variable with index i is modified in statement s; otherwise, the i’th bit is 0. Similarly, you will have a bit vector for relationship Uses (s, v).

### 13.8 Processing PQL queries

Query processing sub-system consists of a query pre-processor, query evaluator and query result projector. Query pre-processor validates a program query written in PQL and builds an internal data structure, so-called query tree. Query validation includes checking if references to design entities, attributes and relationships are used according with their definition in the program design model. The query pre-processor consults the tables describing a program design model to validate queries. Refer to section [Table-driven technique - query validation example](#) in “Compendium of recommended engineering practices” for the project for further details.

Query evaluator traverses the query tree and consults the PKB in order to evaluate a query. Designing a general query evaluation mechanism that can answer any query described in the last section is a complicated task. The task becomes even more complex if you want to

optimize query evaluation with respect to time and/or memory usage. In this project, you will design an evaluator only for certain types of program queries. If you try to implement some optimizations, your project will qualify for a higher grade.

Query result projector formats the query results in the format suitable for displaying.

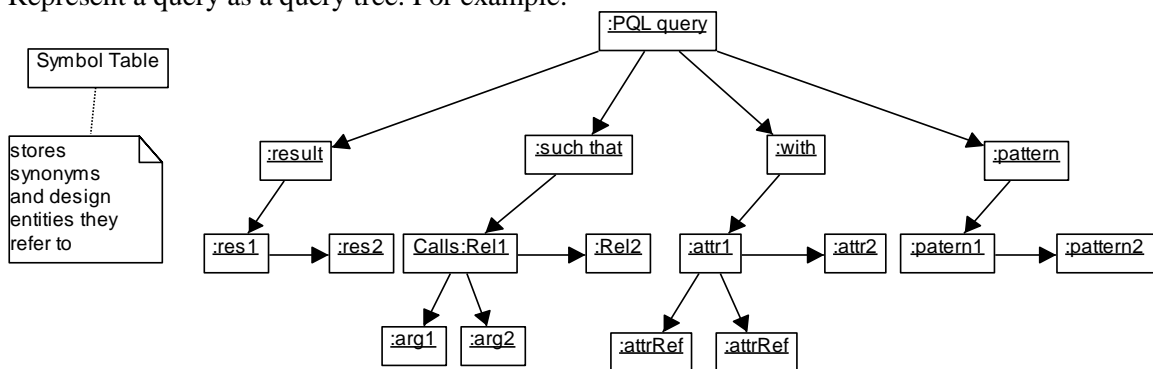
### 13.8.1 Entering program queries (Group-PQL)

You can apply the technique described in section [Parsing SIMPLE](#) to parse PQL queries.

### 13.8.2 Query pre-processor

The role of the query pre-processor is to validate a query and build query tree. Query validation means checking if all the references to program design entities, their attributes and relationships in a program query agree with the program design model definition. Refer to [Table-driven technique - query validation example](#) for further details and design guidelines.

You need represent program queries in the form that is convenient for query evaluation (and eventually optimization). Textual representation is not suitable for that purpose (why?). Represent a query as a query tree. For example:



**Figure 22. A query tree**

The overall structure of a query is represented in a tree form. Logic expressions in a query refer to design entities, their attributes and entity relationships of conceptual models of program design. The query tree can be traversed to evaluate a query.

Make your query pre-processor build the query tree. References to design entities, relationships and attributes in the query tree should be represented by references to entries in the program design model tables used by query pre-processor to validate queries.

### 13.8.3 Query evaluation and optimization

Query evaluator may evaluate queries incrementally as follows. Having built a query tree, interpret query fragments contained in the **such that**, **with** and **pattern** clauses one by one. In that way, you divide the query evaluation process into a sequence of simpler evaluation steps. Evaluation of each query fragment involves calls to one or more interface operations to access information from the PKB. Evaluation of each query fragment produces an intermediate result that approximates the query result.

Here are some basic hints for implementing query evaluator and optimization:

First design the best possible Basic Query Evaluator (BQE) that can correctly evaluate any query without optimizations. Query involving many relationships must be evaluated incrementally, in steps. In the design of BQE pay attention to computing and representing intermediate query results. BQE should be prepared to employ various optimization strategies such as re-arranging the order in which to evaluate relationships in a query and many others. You will have to experiment with various optimization strategies that's why it is important that different optimization strategies can be plugged-into your BQE. BQE should not depend on any specific optimization strategy, but easily work with any optimization strategy you can think of.

The incremental strategy is relatively straightforward to implement but, without optimizations, query evaluation will be slow and will take much memory.

Of course, the size of the source program affects the efficiency of query evaluation. However, design decisions considering the PKB and query evaluator also have much to do with it.

The choice of data representation in the PKB affects performance of query evaluation. If you design your data structures (for AST, CFG, etc.) taking into account the ways you will need to compute information during query evaluation, your query evaluator will run faster. Let us say you have stored only forward links among your CFG nodes. Evaluation of the query that requires traversing the CFG backwards will be very time consuming. The way you store information about variables modified/used will have much impact on the evaluation time of queries that involve relationship Affects().

The main factors that have to do with effective query evaluation is the size of the intermediate result and the evaluation time. Optimizations should reduce the size of the intermediate result and/or the evaluation time. Changing the order in which you evaluate relationships in a query, as well as extra information about the source program may help you evaluate queries in more effective way. Assess various optimizations analytically before you implement them.

Here are samples of query evaluation performance measurements.

assign a, a1, a2, a3;

while w;

procedure p1, p2;

variable v;

- Q1. **Select** <a.stmt#, w.stmt# > **such that** Follows (a, w)
- Q2. **Select** <a.stmt# > **such that** Modifies (p1, a) **and** Calls (p1, "p")
- Q3. **Select** a **pattern** a ("X", \_) **such that** Follows (a, w)
- Q4. **Select** a **such that** Follows (a, w) **and** Next (a, a1)
- Q5. **Select** a **such that** Next (a1, w) **and** Parent (w, a)
- Q6. **Select** <a1, a3> **such that** Follows (a1, a2) **and** Follows (a2, a3)

query	query evaluation time for 7 KLOC source	query evaluation time for 50 KLOC source
Q1	0.029	0.265
Q2	0.875	2.56
Q3	0.053	0.47
Q4	0.05	1.06
Q5	1.1	7.7
Q6	16.02	929

**Table 5. Performance of query evaluator**

Table 5 depicts the system time taken to evaluate queries for 7KLOC and 50KLOC source programs on the Ultra 1 200MHz computer with 256MB memory, running Solaris 2.5.1 (this performance test was done long ago). Traversing abstract syntax trees and looking for syntactic patterns is fast. Consulting the tables to fetch the global design only marginally affects the performance. The reason why the query evaluation time grows rapidly for queries Q5 and Q6 is the computation of joins for large sets of tuples.

*PQL* allows us to constrain syntactic patterns with conditions placed in **such that** clauses. For example, we may search for 'while' statements that contain calls to procedure 'p' and, at the same time, appear on a certain control flow path. This feature gives *PQL* the necessary expressive power but it also poses problems for efficient query evaluation. Firstly, evaluation of queries may require multiple traversals of abstract syntax trees, control flow graphs and other tree-like structures. Secondly, a query evaluator may produce large sets of tuples at intermediate steps of the query evaluation process. If subsequently we perform join operations

on large sets of tuples, then query evaluation will take much computer memory and time. This issue affects the evaluation time for queries Q5 and Q6 (see Table 5).

Various optimizing strategies may offer possible remedies to the above problems. Notice that you may arbitrary change the order in which you evaluate search conditions without changing the query result. But both evaluation time and memory requirements may drastically change from one evaluation strategy to another. By rearranging the order of search conditions in a query, we may reduce time and memory needed for query evaluation. For example, it will take much longer to evaluate query:

```
assign a1; a2; while w;
```

```
Select a1 such that Follows (a1, a2) and Follows (a2, w)
```

than an equivalent query:

```
Select a1 such that Follows (a2, w) and Follows (a1, a2)
```

Generally, we should try to rearrange conditions so that we compute the most restrictive conditions first. By following this rule, intermediate sets of tuples are smaller and the query evaluation time is minimized.

Analysis of arguments of relationships can be also helpful. Consider the following query:

```
assign a; while w;
```

```
Select a such that Follow (a, w)
```

To evaluate this query, we could start by finding either assignments ('assign') or while loops ('while'). If a program contains more assignments than while loops (most programs do), the query will evaluate faster if we find while loops first. Therefore, an optimized query evaluator could attempt to find while loops first and then compute assignments such that Follow (assign, while).

The above are examples of issues that you can take into account when optimizing queries.

## References

Fowler, M. *UML Distilled, Second Edition*, Addison-Wesley, 2000

Meyer, B. *Software Construction, Second Edition*, Prentice Hall, 1997, QA76.64 Mey (pp. 2-19, 39-64)

## Appendix A. Summary of PQL grammar rules

Lexical tokens are written in capital letters (e.g., LETTER, INTEGER, STRING). Keywords are between apostrophes (e.g., 'procedure'). Non-terminals are in small letters.

*Meta symbols:*

a\* - repetition 0 or more times of a

a+ - repetition 1 or more times of a

a | b - a or b

brackets ( and ) are used for grouping

*Lexical rules:*

LETTER : A-Z | a-z -- capital or small letter

DIGIT : 0-9

IDENT : LETTER ( LETTER | DIGIT | '#' )\*

INTEGER : DIGIT+

*Auxiliary grammar rules:*

tuple : elem | '<' elem ( ',' elem )\* '>'

elem : synonym | attrRef

synonym : IDENT

attrName : 'procName' | 'varName' | 'value' | 'stmt#'

entRef : synonym | '\_' | ''' IDENT ''' | INTEGER

stmtRef : synonym | '\_' | INTEGER

lineRef : synonym | '\_' | INTEGER

design-entity : 'procedure' | 'stmtLst' | 'stmt' | 'assign' | 'call' | 'while' | 'if' | 'variable' |



‘constant’ | ‘prog\_line’

*Grammar rules for select clause:*

select-cl : declaration\* **Select** result-cl ( with-cl | suchthat-cl | pattern-cl )\*

declaration : design-entity synonym ( ‘,’ synonym)\* ‘;’

result-cl : tuple | ‘BOOLEAN’

with-cl : **with** attrCond

suchthat-cl : **such that** relCond

pattern-cl : **pattern** patternCond

attrCond : attrCompare ( **and** attrCompare )\*

attrCompare : attrRef ‘=’ ref

attrRef : synonym ‘.’ attrName

ref : ‘’’ IDENT ‘’’ | INTEGER | attrRef

relCond : relRef ( **and** relRef )\*

relRef : ModifiesP | ModifiesS | UsesP | UsesS | Calls | CallsT |

Parent | ParentT | Follows | FollowsT | Next | NextT | Affects | AffectsT

ModifiesP : ‘Modifies’ ( ‘(’ entRef ‘,’ entRef ‘)’

ModifiesS : ‘Modifies’ ( ‘(’ stmtRef ‘,’ entRef ‘)’

UsesP : ‘Uses’ ( ‘(’ entRef ‘,’ entRef ‘)’

UsesS : ‘Uses’ ( ‘(’ stmtRef ‘,’ entRef ‘)’

Calls : ‘Calls’ ( ‘(’ entRef ‘,’ entRef ‘)’

CallsT : ‘Calls\*’ ( ‘(’ entRef ‘,’ entRef ‘)’

Parent : ‘Parent’ ( ‘(’ stmtRef ‘,’ stmtRef ‘)’

ParentT : ‘Parent\*’ ( ‘(’ stmtRef ‘,’ stmtRef ‘)’

Follows : ‘Follows’ ( ‘(’ stmtRef ‘,’ stmtRef ‘)’

FollowsT : ‘Follows\*’ ( ‘(’ stmtRef ‘,’ stmtRef ‘)’

Next : ‘Next’ ( ‘(’ lineRef ‘,’ lineRef ‘)’

NextT : ‘Next\*’ ( ‘(’ lineRef ‘,’ lineRef ‘)’

Affects : ‘Affects’ ( ‘(’ stmtRef ‘,’ stmtRef ‘)’

AffectsT : ‘Affects\*’ ( ‘(’ stmtRef ‘,’ stmtRef ‘)’

patternCond : pattern ( **and** pattern )\*

pattern : assign | while | if

assign : synonym ( ‘(’ entRef ‘,’ sub-expression-spec | ‘\_’ ‘)’

sub-expression-spec : sub-expression

| ‘\_’ sub-expression

| sub-expression ‘\_’

| ‘\_’ sub-expression ‘\_’

sub-expression : ‘’’ sub\_expr ‘’’

// ‘synonym’ above must be of type ‘assign’

// sub\_expr must be a well-formed expression in SIMPLE, except that brackets are not used

// see query examples in section 7.5 in the Project Handbook

if : synonym ( ‘(’ entRef ‘,’ ‘\_’ ‘,’ ‘\_’ ‘)’

// ‘synonym’ above must be of type ‘if’

while : synonym ( ‘(’ entRef ‘,’ ‘\_’ ‘)’

// ‘synonym’ above must be of type ‘while’

**Summary of other PQL rules:**

1. PQL query can contain any number of **such that**, **with** and **pattern** clauses. All the clauses may appear more than one time in a query, in any order, e.g.:

**Select ... with ... such that ... with ... with ... pattern ... such that ...**

2. There is a default **and** between any two consecutive clauses. Therefore, we can swap or merge clauses without changing the meaning of the query.
3. A query result must satisfy all the query conditions in all the clauses. The existential quantifier is always implicit in *PQL* queries. That means, a query returns any result for which **THERE EXISTS** a combination of synonym instances satisfying the conditions specified in all the query conditions.
4. If the query result clause is a tuple, then tuple elements that satisfy all the query conditions at the same time are reported as query result.
5. Query with result **BOOLEAN** returns true iff there is at least one combination of synonyms satisfying all the query conditions.
6. All the synonyms used in a query must be declared.
7. Spaces can be freely used and are meaningless (also multiple spaces).
8. Arguments in relationships should be synonyms, ‘\_’ and, depending on the relationship, integer (statement line numbers or program line numbers) or string (variable or procedure names). Relationship arguments should conform to program design abstractions models for **SIMPLE**, as defined in Section 6.
9. Underscore ‘\_’ is a placeholder for an unconstrained synonym. Symbol ‘\_’ can be only used when the context uniquely implies the type of the design entity denoted by ‘\_’.
10. Under **with** we can compare an attribute value and constant (integer or string, depending on the type of attribute) or two attribute values (provided they are of the same type).
11. Please check conventions described in Section 7.2, as they apply in addition to the above *PQL* core rules.
12. You can make your own assumptions about any other details that have not been specified in the Handbook – such as case-sensitivity of identifiers/keywords.

--- The End ---