Proper Placement of Derived Classes in the Class Hierarchy

Reda Alhajj
University of Calgary, Canada

Faruk Polat
Middle East Technical University, Turkey

INTRODUCTION

Users may derive new classes by defining views based on the current database contents. Some virtual classes are classified as brothers of existing classes, and others are either superclasses or subclasses of existing base and virtual classes. A base class is defined directly by the user using class definition constructs. A virtual class is classified as a brother of another class if it is derived from the latter class via a selection. To have a homogeneous system, virtual classes must be treated as first-class citizens in an object-oriented model.

In this article, we handle reusability maximization by investigating the proper location of a derived class in the hierarchy by adjusting the list of superclasses and the set of subclasses for a given class to increase inheritance instead of duplication. Our aim is to maximize inherited, and minimize locally defined characteristics by adjusting the list of superclasses to include classes that maximize inherited and minimize locally defined characteristics.

Depending on the specification of their superclasses and subclasses during query evaluation, virtual classes may be classified into four groups: (1) classes for which both superclasses and subclasses are specified; (2) classes for which only superclasses are specified; (3) classes for which only subclasses are specified; and (4) classes for which neither superclasses nor subclasses are specified. Reusability decreases from the first group to the fourth group. Two algorithms are presented to achieve the target of maximizing reusability when possible.

BACKGROUND

Shown in Figure 1 is a class hierarchy that will be referenced frequently in the article where illustrating examples are necessary. Given next is the basic terminology necessary to understand the rest of this article:

1. \( P_d(c) \): set of classes used in deriving class \( c \); \( P_d \) is empty for base classes because they are user-defined.
2. \( C_p(c) \): list of direct superclasses of class \( c \).
3. \( C_b(c) \): set of direct subclasses of class \( c \).
4. \( L_{\text{instances}}(c) \): set of object identifiers of objects added directly to class \( c \).
5. \( W_{\text{instances}}(c) \): set of object identities of objects in class \( c \) and its direct and indirect subclasses.
6. \( W_{\text{attributes}}(c) \): set of attributes that determine the state of objects in \( L_{\text{instances}}(c) \). It includes inherited and locally defined attributes.
7. \( L_{\text{attributes}}(c) \): set of additional attributes locally defined in class \( c \), that is, \( L_{\text{attributes}}(c) \subseteq W_{\text{attributes}}(c) \).
8. \( W_{\text{behavior}}(c) \): set of methods common to all objects in class \( c \). It includes both inherited methods and locally defined methods.
9. \( L_{\text{behavior}}(c) \): set of additional methods locally defined in class \( c \), that is, \( L_{\text{behavior}}(c) \subseteq W_{\text{behavior}}(c) \). For every virtual class \( c \), \( L_{\text{behavior}}(c) \) includes a method, called FindObjects(), which implements the query expression that decides on objects in \( W_{\text{instances}}(c) \).
10. \( m(c, f, r) \): the header of method \( m \) such that, \( m \in L_{\text{behavior}}(c) \), \( f \) is the list of domains for formal parameters, and \( r \) is the domain of the result.
Example 1: [Base classes] All the classes shown in Figure 1 are base classes. Next are the characteristics of some of them, where B, J, P, SI, RP, R, JP, and CP stand for Book, Journal, Proceedings, SpecialIssue, ResearchPaper, Researcher, JournalPaper, and ConferencePaper, respectively:

C\(_i\)(B)=\{OBJECT\} \quad C\(_j\)(B)=\{P\} \quad P\(_i\)(B)=\phi

L_{attributes}(B)=\{title:String, author:[R], year:Date, subject:String, publisher:String\}
L_{behavior}(B)=\{title(), title(t), author(), author(p), year(), year(y), subject(), subject(s), publisher(), publisher(p)\}

C\(_i\)(P)=\{B\} \quad C\(_j\)(P)=\{SI\} \quad P\(_j\)(P)=\phi

L_{attributes}(P)=\{location:String, chairperson:R, committee:{R}, AcceptanceRate:Real, contents:[CP]\}
L_{behavior}(P)=\{location(), location(l), chairperson(), chairperson(p), committee(), committee(p), AcceptanceRate(), AcceptanceRate(i), contents(), contents(c)\}

C\(_i\)(J)=\{OBJECT\} \quad C\(_j\)(J)=\{SI\} \quad P\(_j\)(J)=\phi

L_{attributes}(J)=\{title:String, EditorInChief:R, EditorialBoard:{R}, subject:String, publisher:String, contents:[JP]\}
L_{behavior}(J)=\{title(), title(t), EditorInChief(), EditorInChief(e), EditorialBoard(), EditorialBoard(p), subject(), subject(s), publisher(), publisher(p), contents(), contents(c)\}

Example 2: [Brother class] Given DBJournals as a brother of class J with: P\(_j\)(DBJournals)=\{J\}. By definition, DBJournals has the following characteristics:

C\(_p\)(DBJournals)=\{\} \quad C\(_m\)(DBJournals)=\{\}
L_{attributes}(DBJournals)=\{\} \quad L_{behavior}(DBJournals)=\{\} \quad W_{instances}(DBJournals)=\{\}

L_{attributes}(J)=\{title:String, EditorInChief:R, EditorialBoard:{R}, subject:String, publisher:String, contents:[JP]\}
L_{behavior}(J)=\{title(), title(t), EditorInChief(), EditorInChief(e), EditorialBoard(), EditorialBoard(p), subject(), subject(s), publisher(), publisher(p), contents(), contents(c)\}

Example 3: [Virtual classes] Consider the following characteristics of some example derived classes.

PP: (PrintedPublications) includes all books and journals.

P\(_i\)(PP)=\{B, J\} \quad C\(_i\)(PP)=\{OBJECT\} \quad C\(_j\)(PP)=\{B, J\}
W_{instances}(PP)=W_{instances}(B) \cup W_{instances}(J) \quad L_{instances}(PP)=\{\}
L_{attributes}(PP)=\{title:String, subject:String, publisher:String\}
L_{behavior}(PP)=\{title(), title(t), subject(), subject(s), publisher(), publisher(p), FindObjects()\}

NP: (NewProceedings) includes only some characteristics from class P.

P\(_i\)(NP)=\{P\} \quad C\(_i\)(NP)=\{OBJECT\} \quad C\(_j\)(NP)=\{P\}
W_{instances}(NP)=W_{instances}(P) \quad L_{instances}(NP)=\{\}

L_{attributes}(NP)=W_{attributes}(B) \cup \{location:.String, chairperson:R\}
L_{behavior}(NP)=\{W_{behavior}(B) \cup \{location(), location(l), chairperson(), chairperson(p), FindObjects()\}\}

MAIN THRUST

Locally defined characteristics of class c include L_{attributes}(c) and L_{behavior}(c). The inherited characteristics of class c are (W_{attributes}(c)–L_{attributes}(c) and (W_{behavior}(c)–L_{behavior}(c)). In this section, we present two algorithms to investigate adjusting for a given class: (1) its list of superclasses and (2) its set of subclasses. The first is necessary for classes in the third group, and the second is suitable for classes in the second group. However, both algorithms are required for classes in the fourth group.

Algorithm 1 investigates the possibility of maximizing reusability by adding some existing classes to C\(_i\)(c) or by pushing class c into the list of superclasses of some existing classes. We employ the following heuristic; having C\(_b\)(c) not empty helps in limiting our search to the direct and indirect superclasses of every class in C\(_b\)(c). Otherwise, we have to consider and investigate all classes in the hierarchy, which is the general case when C\(_b\)(c) is empty. For each class c\(_p\)\in C\(_b\)(c), Definition 1 is used to find the inheritance list of class c\(_p\), denoted IL(c\(_p\)), that includes all its direct and indirect superclasses. For the other case of having C\(_b\)(c) empty, the class hierarchy must be traced top-down and left to right in order to include all the classes present in the hierarchy inside IL(c). Algorithm 1 checks for the possibility of including inside C\(_p\)(c) any of the classes in IL(c\(_p\)), and the list of superclasses of the considered class is adjusted accordingly.

Definition 1: [Inheritance List] Given any class c, IL(c) is defined using:

1. For every c\(_p\)\in C\(_b\)(c), c\(_p\)\in IL(c);
2. For every c\(_p\)\in IL(c), classes in C\(_p\)(c\(_p\)) are placed at the head of IL(c);
3. Nothing else belongs to IL(c).

Example 4: [Inheritance List] IL(SI)=\{B, J, P\} and IL(P)=\{B\}. 

487