Uniform Comparison of Data Models Using Containment Modeling

Jim Whitehead
ejw@cs.ucsc.edu
Containers

- Our daily lives are filled with containers
  - Backpacks, suitcases, purses, wallets, pockets, cups, bottles, cans, baskets, boxes, cabinets, drawers, closets, bookshelves
  - Items are **physically contained** within these containers
  - Each item can only belong to one container at a time
  - Of course, nesting is allowed: bottle in suitcase in trunk in car
Digital Objects are Different

- They can be duplicated at low to trivial cost
  - Can be used non-destructively by several people at the same time
- Assignment and management of object identifiers can be automated
- Dramatically increases the utility of containing digital objects using references
Insights

- The data models of hypertext systems can be viewed as networks of containment relationships.
  - Composites, links, anchors, objects … all are containers
- It is valuable to have a model focused on just the containment relationships among hypertext system entities.
  - Other models, such as UML structure and sequence diagrams, are complementary
Goals of the Modeling Mechanism

- **Uniformity**
  - Use a minimal set of system-neutral abstractions

- **Utility**
  - Answer basic questions about data models:
    - Are links to whole works, or to a subregion within a work?
    - Are links separate from, or part of a work?

- **Support Analysis**
  - Tease out design choices

- **Graphic Formalism**
  - Communicate models quickly and intuitively with figures

- **Concise Format**
  - Fit many models on a single page, for easy comparison

- **Cross-discipline**
  - Capable of modeling other content management systems
Containment Data Modeling

- A constrained form of extended entity-relationship modeling
- Entities are hypertext system abstractions
  - Use abstractions and terminology defined by each system
- Relationships are one of
  - **Containment**
    - Inclusion or referential
  - **Storage**
    - A distinguished form of inclusion containment across the storage system (filesystem) to hypertext system boundary
  - **Inheritance**
    - Used *only* to reduce clutter
    - Children inherit all relationships of the parent
    - Create a separate, complementary UML structure diagram for the complete inheritance picture
Basic Static Containment

- **Container**: a set of entities
- **Aspects of containment model**:
  - Abstract properties
    - Mathematic set properties independent of a specific computer representation
  - **Containment**: object belongs to one set (single) or many (multiple)
  - **Membership**: each object can belong to a set once (single), or many (multiple) times
  - **Ordering**: persistently ordered, unordered, indexed, grouped
- **Containment type**
  - How the computer represents that a container contains a particular entity
    - **Referential** - a pointer to another object
    - **Inclusion** - contained object is a sub-part of container object
Containment Types

- **Inclusion**: item is physically part of container
- **Referential containment** types:
  - Describes *identifier storage*
    - On container
    - On object
    - On container and object
  - First-class containment relationship
    - A new container is introduced, and storage is delegated to it.
  - Hybrid
    - First-class container plus one other
Dynamic Containment

- Queries can be used to determine collection membership
  - Query results specify the endpoint of a single containment relationship
    - Often used in hypertext versioning, where the query specifies a revision selection criteria within a versioned object
    - Query typically returns a single object
    - Storage of query predicate varies
  - Query results specify the membership of the collection
    - Query returns multiple objects
Three-layer model of containers

Abstract Relationship Layer

Relationship Layer

Explicit Relationship Layer

Example #1: Inclusion

Contains - single containment, single membership, unordered, inclusion

Example #2: Referential

Contains - multiple containment, single membership, ordered, containment relationship on container

Concrete Representation Layer

File with a linked list of content chunks

Container data item uses linked list of identifiers of member data items

* = has identifier

container ➔ contained entity

container* ➔ contained entity

container* ➔ contained*

id0

id1 id2 id3

data item id1

data item id2

data item id3
Graphical Notation

Inclusion and referential containment relationships can be decorated with multiplicity information (1:N, M:N, etc.)

**Common Relationships:**

- **Containment (unordered inclusion)**
  - inclusion, single containment, single membership, unordered

- **Containment (ordered inclusion)**
  - inclusion, single containment, single membership, ordered

- **Containment (referential, unordered)**
  - referential, multiple containment, single membership, unordered

- **Containment (referential, ordered)**
  - referential, multiple containment, single membership, ordered

- **Containment (referential, multiple membership, ordered)**
  - referential, multiple containment, multiple membership, ordered
Dexter

- **Composite**
  - Link
  - Endpoint specification
  - Anchor
  - Content
  - Attribute
  - Presentation specification

- **Inheritance**
  - Single containment, single membership, unordered, inclusion
  - Multiple containment, single membership, unordered, referential
  - Single containment, single membership, ordered, inclusion
Observations on the Dexter Model

- Concisely represents the Dexter data model
- Development required many hours of careful study of CACM’94 article, and Dexter Z specification
  - Now you can understand the containment model in much less time
- I got it wrong at first
  - Figure 8 of my PhD dissertation 😊
  - But, was later able to update my understanding

- Containment modeling is a useful way to capture, reflect upon, and analyze system data models.
The Web

Web server

resource

HTML link

Stores

Multiple containment, single membership, unordered

Single containment, single membership, unordered, inclusion

Single containment, single membership, ordered, inclusion
Observations

- HTTP and URLs (vs NFS and filenames) are the big difference

- Pattern for embedded linking

Stores

- Single containment, single membership, unordered, inclusion

Multiple containment, single membership, unordered

Single containment, ordered, inclusion
NoteCards

filesystem

notefile

notecard

browser  filebox  sketch  tabletop

Inheritance

Stores

Multiple containment, single membership, unordered

Single containment, single membership, unordered, inclusion
Observations

- Pattern for linked card spaces with links not contained by card

Inheritance

Stores

Multiple containment, single membership, unordered
Intermedia

Stores

Multiple containment, single membership, unordered

Single containment, single membership, unordered, inclusion
Sun’s Link Service

- **filesystem**
- **link database**
  - **link**: N
  - **object**: M
  - **file**: 1

No anchoring, not entirely comparable

Stores

- **Single** containment, single membership, unordered, inclusion

Multiple containment, single membership, unordered
Pattern for link databases/Webs separable from linked documents

Stores

Multiple containment, single membership, unordered

Single containment, single membership, unordered, inclusion
HURL (SP3/HB3)

- One of the most complex versioning unaware hypertext data models.
- David Hicks added versioning to this!

**Multiple** containment, single membership, unordered
Contributions

- Containment data modeling is capable of uniformly and concisely modeling a wide range of existing data models
  - 14 in the paper (22 systems total, including other work)
- Permits structural cross-comparison of system data models
  - Patterns jump out
  - Can identify design choices and spaces
- Dexter and the Web can be modeled, just like other hypertext systems
  - Embody specific design tradeoffs
Future Work

- Apply containment modeling to a broad range of content management systems
  - Software Configuration Management
  - Document Management
  - VLSI CAD
  - Goal: a deep understanding of the design spaces of content management

- Improve containment modeling formalism
  - Better representation of constraints
  - Addition of AND/OR nodes
  - Improved graphic representation (better orthogonality)

- Automatic generation of hypertext, and content management repositories
  - Specification → generator → repository

- Working with students Dorrit Gordon, Guozheng Ge
Knowledge of Containment Starts Young!