Guava: Capturing the Intrinsic Organization of Knowledge in User Interfaces

James Terwilliger and Lois Delcambre

Computer Science Department
Portland State University

Judith Logan, MD

Division of Medical Informatics & Clinical Epidemiology

Oregon Health & Science University

Clinical Outcomes Research Initiative (CORI)

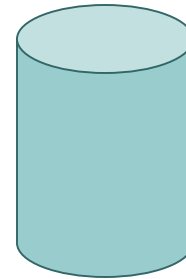DB structure is typically "generic" with no indication of which table is which

UI
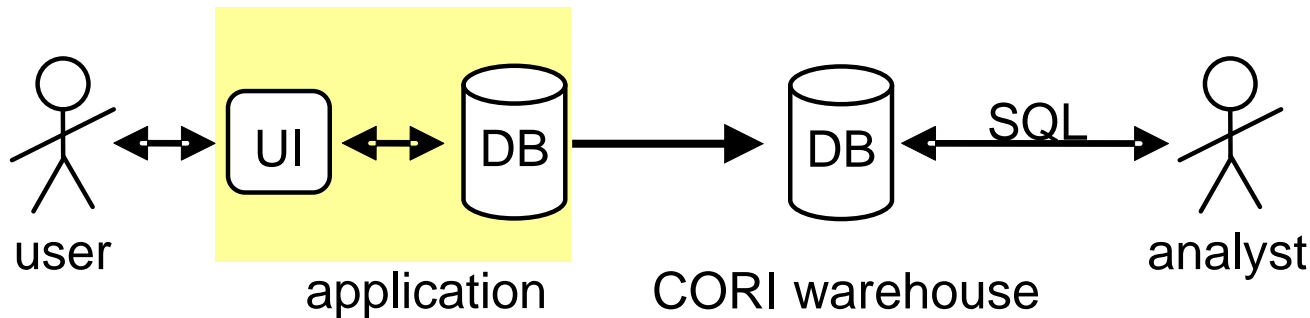report

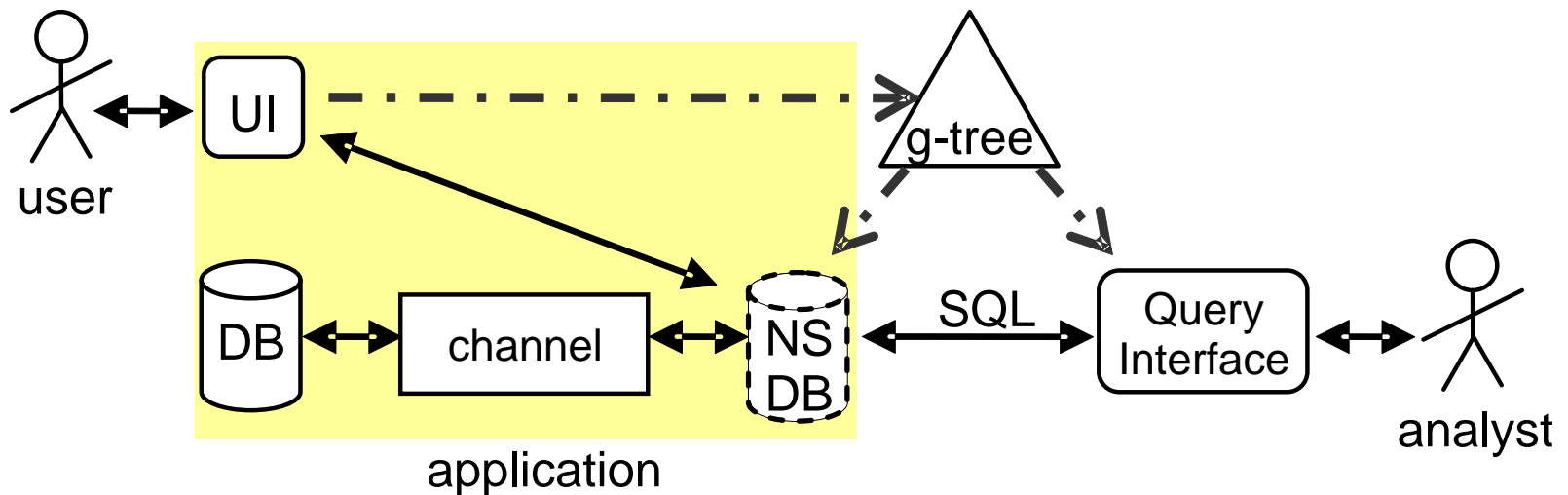de-identified patient medical reports

CORI Warehouse

extracted data

CORI Analysts

statistical analysis to study and improve the practice of endsocopy

UI
report

The Traditional Approach vs. The Guava Approach

Traditional approach: analyst writes queries against (physical) DB

GUAVA: UI generates g-tree, then g-tree generates natural schema

Problem Statement

- The data analysts at CORI are experts in statistical method and clinical terminology

- They are not necessarily database or programming experts

- The only knowledge organization systems available to them are database schemas

To Make Matters More Difficult…

- In the past, there was only one source of data

- Soon, they may be analyzing data from as many as five, each with its own arcane schema

Guava

- Builds an ontology DIRECTLY from the user interface for the reporting tool – one for each data source/UI

- Use this ontology (from the UI) as a query interface

- All UI information is now also searchable

A Simple UI and its Implied Ontology (Guava Tree)
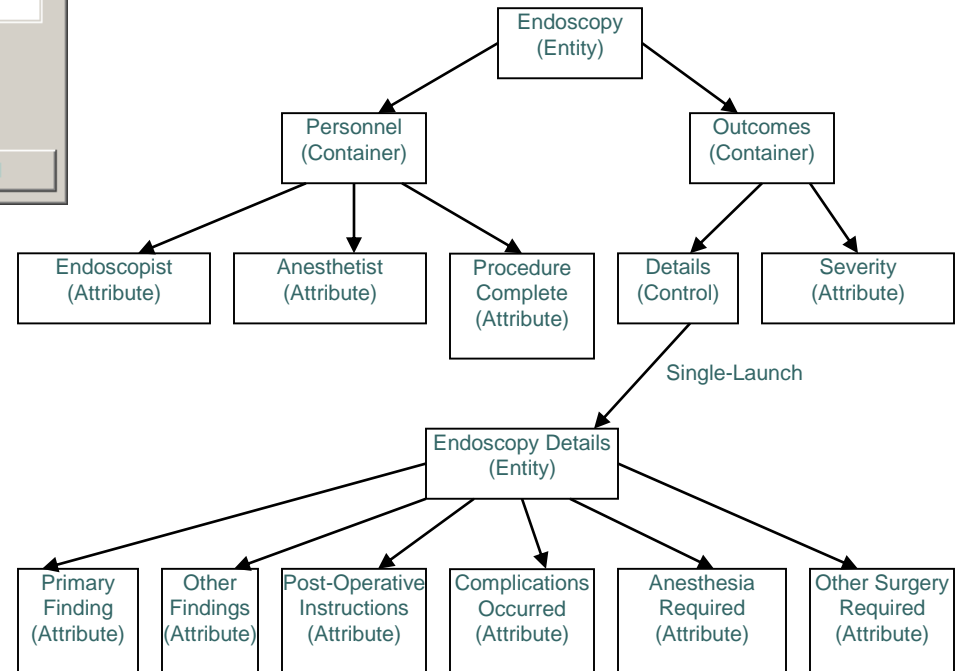
Simple Query Against a Guava Tree

Bold means "print"; filters shown inline

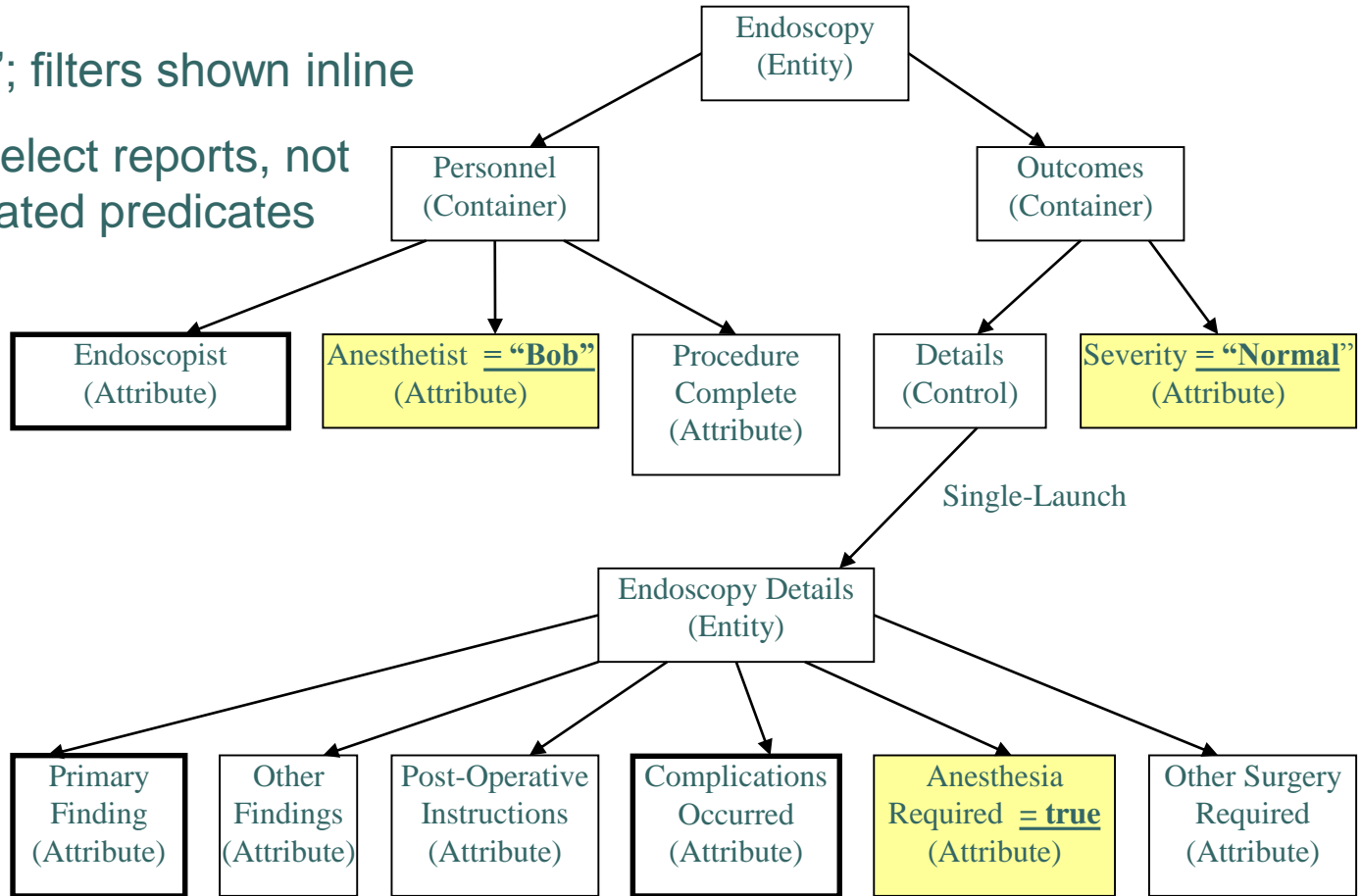Query is used to select reports, not evaluate sophisticated predicates or calculations

Classifiers

- Allows user to conform the elements and domains of one Guava Tree to those of another

Classifier Habits (Cancer)
  Classifies packs per day according to conversations with cancer study on 5/3/02

None	$\leftarrow$	$PacksPerDay = 0$
Light	$\leftarrow$	$0 < PacksPerDay < 2$
Moderate	$\leftarrow$	$2 \leq PacksPerDay < 5$
Heavy	$\leftarrow$	$PacksPerDay \geq 5$

Classifier Habits (Chemistry)
  Classifies packs per day according to flier from chemical studies

None	$\leftarrow$	$PacksPerDay = 0$
Light	$\leftarrow$	$0 < PacksPerDay < 1$
Moderate	$\leftarrow$	$1 \leq PacksPerDay < 2$
Heavy	$\leftarrow$	$PacksPerDay \geq 2$

Classifier Tumor Size
  Estimates tumor volume based on dimensions in 3-space. Assumes 52% occupancy from sphere-to-cube ratio.

$TumorX * TumorY * TumorZ * 0.52$ $\leftarrow$ $TumorX > 0$ AND $TumorY > 0$ AND $TumorZ > 0$

Analyst Feedback (Informal)

- Held on September 14, 2006
- Demonstrated query interface capabilities of early prototype to the CORI analysts
- Response was entirely positive
  - "So much potential"
  - "Very useful"
  - "Exciting"
- Most excited about the capability of searching the content of the UI

Status

- Early prototype is complete, showing the Guava Tree as a tree structure
- Next version will use mock-ups of UI
  - Pose queries by entering sample data in form
  - Returns results that match the sample data
  - View results in context of the form through which it was entered