Who won the Universal Relation wars?

Alberto Mendelzon
University of Toronto

Who won the Universal Relation wars?

Who won the Universal Relation wars?

The Good Guys.

The Good Guys

Outline

- The UR assumption in schema design
- The UR as UI
- Information integration and the UR
- Losslessness and Views
- Data Exchange and the UR
- Practical impact
- Conclusions

In the beginning: schema synthesis

[Delobel & Casey, 1972] Decomposition method

[Wang & Wedekind, 1975] Informal synthesis method

[Bernstein, 1976] 3NF Synthesis

- Start with a set of time-varying functions. Produce a set of 3NF relation schemes that embody the functions in their keys.
- Uniqueness assumption: "...if there are two fd's on the same set of attributes, then they are the same fd."
- Universal Relation assumption [Beeri, Bernstein, Goodman 1978]: "...all relations in a database are projections of a single relation."

UR Wars I

[Beeri, Bernstein & Goodman 1978] *"This 'universal relation assumption' is a controversial issue in the field."*

[Bernstein & Goodman 1980] *"Normalization theory will remain an isolated theoretical area, divorced from database practice, until this assumption is circumvented."*

[Kent 1982] *"The UR is an unsatisfactory model for relational theory or practice."*

[Ullman 1983] *"...the UR asumption here is like the wave theory of light; it provides lots of insights and fosters a useful design methodology, even though there may be some technical problems with [it] when applied in the wrong context."*

Logical data independence

- Physical data independence: specify a query, not the physical access paths to compute it.
- Logical data independence: specify a set of attributes, not the logical navigation path to compute the connection among them.
- [Carlson & Kaplan 1976, Osborn 1979]: Algorithms for inferring logical access paths to arbitrary sets of attributes.

The UR as UI

[Fagin, Honeyman, Maier, Sagiv, Ullman, Vardi, Yannakakis, etc.]:

- For each database state s consistent with dependencies D there is at least one universal relation u satisfying D such that $s(R_i) \subseteq \pi_{R_i}(u)$. (weak UR assumption)
- The connection or window $[X]$ on a set of attributes X on state s is the set of X-tuples that belong to every universal relation u for s.
- Canonical way of computing $[X]$: build a representative instance for s that is homomorphic to every possible u, project on X.

UR Wars II

[Atzeni & Parker 1982] *"Except in small databases or in views, we are afraid that users will misinterpret the intention behind the universal schema."*

[Codd 1988] *"I believe the universal-relation model fails completely as an alternative to the relational model."*

UR Wars II

[Atzeni & Parker 1982] *"Except in small databases or in views, we are afraid that users will misinterpret the intention behind the universal schema."*

[Codd 1988] *"I believe the universal-relation model fails completely as an alternative to the relational model."* (in response to [Vardi 1988]: *"The universal relation model is not meant to replace the relational model; rather, it is meant to supplement it."*)

UR Wars II

[Atzeni & Parker 1982] *"Except in small databases or in views, we are afraid that users will misinterpret the intention behind the universal schema."*

[Codd 1988] *"I believe the universal-relation model fails completely as an alternative to the relational model."* (in response to [Vardi 1988]: *"The universal relation model is not meant to replace the relational model; rather, it is meant to supplement it."*)

[Ullman 1982] "The fact that systems using the concept exist and are successful, even in a limited context, is a more powerful refutation of the argument against the UR idea than are any convolutions of logic that the author can provide."

Information integration and the UR

[Maier, Ullman, Vardi 1984]

"Suppose we have a universal relation scheme ABC with relations [(COURSE,STUDENT), (STUDENT,ENROLLMENT), and (COURSE,ENROLLMENT)]. ...we may imagine that a relation (COURSE,STUDENT), representing graduate courses, was at some time merged with a network database representing the many-many relationship between undergraduate courses and students by means of dummy ENROLLMENT records..."

Global and local schemas

Given (virtual) instance on global schema G and (real) instance defined on local schemas $S_1, \ldots, S_k$, compute query Q on G.

What's the connection between the global instance and the local relations?

- LAV [Levy, Rajaraman & Ordille 1996]: each relation in each S_i is defined by a view on G
- GAV [García Molina et al, 1995]: each relation in G is defined by a view on S.

UR and LAV

LAV query semantics: $Q(db)$ is the set of tuples that belong to every instance of G for db. (certain answers)

UR $=$ special case of LAV where G is a single-relation schema.

LAV variants: [Abiteboul & Duschka 1998, Calvanese et al, 2000]

- exact views (CWA): global instances yield exactly the given local relations – pure UR assumption
- sound views (OWA): global instances yields supersets of the given local relations – weak UR assumption

Losslessness

[Rissanen 1977, Aho, Beeri & Ullman 1979] A decomposition of a relation scheme into a set of its projections is lossless if any instance of the relation can be recovered from the projections by the natural join.

Theorem A decomposition is *information preserving* iff it is lossless.

Corollary: to answer a query Q on U given its lossless projections $R_1, \ldots, R_k$ just rewrite Q by substituting $R_1 \bowtie \ldots \bowtie R_k$ for U.

Losslessness and Views

[Calvanese et al 2002] “A set of views is lossless with respect to a query if, no matter what the database is, we can answer the query by solely relying on the content of the views.”

[Bawa, Li, and Ullman 2001] Comparing sets of views wrt the set of queries that they can answer.

Some very basic questions still open, e.g.:

If a set of conjunctive views is lossless wrt a conjunctive query Q, can Q always be rewritten to answer the query?

Data Exchange and the UR

Data Exchange [Popa et al 2002, Fagin et al 2003]:

- Given source schema S, target schema T, source-to-target dependencies, and source instance I, materialize a target instance J of T (a solution) such that I and J satisfy all the constraints.
- A universal solution is one that has homomorphisms to all other solutions.
- To construct a universal solution: start with the pair $(I, \emptyset)$ and chase wrt all the constraints.

Practical importance

- “It is not certain whether ten years from now relational database systems will routinely provide a universal relation interface for those applications that can have one, or whether the idea will turn out to have no practical importance.” [Ullman 1983]
- “Perhaps 10 years from now [the UR model] will be as commercially successful as the relational model is today.” [Vardi 1988]

Practical importance

- The idea seems to be rediscovered every few years [Many rejected papers, 198X-20XX]
- Sabrix: While traditional audit databases are efficient in storing and managing data, they are not user-friendly when it comes to easily identifying related information to query for reporting purposes. The Sabrix System removes this complexity by providing the Universal Relation view, which effectively merges the three tables into a single table that can be queried by a third-party reporting tool...
- Lotus Notes: Upon connection to a Notes database, NotesSQL generates a special table [...] called the Universal Relation. The Universal Relation contains all fields defined in all forms in the Notes database.

Conclusions

- The UR work could use a serious reassessment (this talk is not it).
- *"This talk also argues that most new ideas are not, in fact, new but have been proposed before. As such, it is important to have a strong historical perspective, so we avoid reinventing the wheel and repeating historical mistakes."*

Conclusions

- The UR work could use a serious reassessment (this talk is not it).
- *"This talk also argues that most new ideas are not, in fact, new but have been proposed before. As such, it is important to have a strong historical perspective, so we avoid reinventing the wheel and repeating historical mistakes."* [Stonebraker 2002]