

Checking Entity-Relationship Models Interactively to Improve Database Schema Design

The apparent ease with which Entity-Relationship (ER) models can be understood and created can lead to false impressions of the quality of those models, which can result in non-functional dependencies in a relational schema. Such problems should be resolved at the ER model stage rather than by subsequent normalization.

We have developed an approach to identifying potential [redundancy](#) and [ambiguity](#) within an ER model, and implemented our approach in a prototype CASE tool, ERD2. ERD2 interacts with the modeller to help resolve these problems to improve the quality of the ER model and, hence, of the subsequently derived database schema.

The approach is generally applicable to a range of graphical modelling notations that utilise the same basic constructs.

Researcher

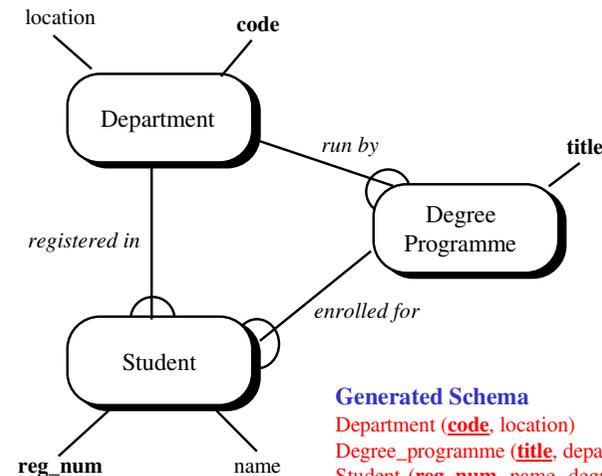
David S. Bowers - Department of Computing, The Open University

Recent Publications

Bowers, D.S (2000). 'Database Schema Improvement Techniques for CASE Tools' - UKAIS 2000.

Bowers, D.S (2002). 'Detection of Redundant Arcs in Conceptual Models' - International Workshop on Conceptual Model Quality, Turku, Finland

Collaborator: Alex Gray, Professor of Advanced Information Systems, Cardiff University



Generated Schema

Department (**code**, location)

Degree_programme (**title**, department_code)

Student (**reg_num**, name, degree_programme_title,

department_code)

Redundant field - transitively dependent on "degree_programme_title"

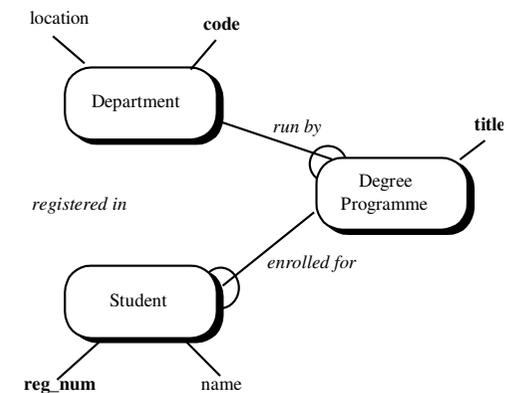
Redundancy

If an ER model contains multiple paths between two entities, one path is redundant if BOTH its signature and its semantics are implied by another

Redundant relationships in the ER model lead to redundancy in a generated relational schema

Ambiguity

If the "registered_in" relationship (path) is removed, but it actually has *different* semantics from those implied by a student being "enrolled for" a degree programme which is "run by" a department, then that remaining path is overloaded and ambiguous



The Trouble with Graphical Languages...

... is that you can draw silly things!

Diagrams such as Entity Relationship (ER) Models are intended for communication between analyst and client. Although they seem to be easy to understand, there is evidence that even *analysts* may not understand them completely.

Furthermore, interaction with clients tends to be confined to the early stages of analysis, whereas questions of potential redundancy are often not resolved until later, perhaps without full knowledge of the client domain

One problem can be the inclusion in the ER model of relationships which appear to be specified in an initial system description, but which are, in fact, implied by other parts of the model. Such relationships are redundant, and, unless they are detected and removed, they will lead to redundancy in any relational schema synthesised from the model.

Researchers

David S. Bowers - Department of Computing, The Open University

Recent Publications

Bowers, D.S (2000). 'Database Schema Improvement Techniques for CASE Tools' - UKAIS 2000.

Bowers, D.S (2002). 'Detection of Redundant Arcs in Conceptual Models' - International Workshop on Conceptual Model Quality, Turku, Finland

Collaborators: Alex Gray, Professor of Advanced Information Systems, Cardiff University

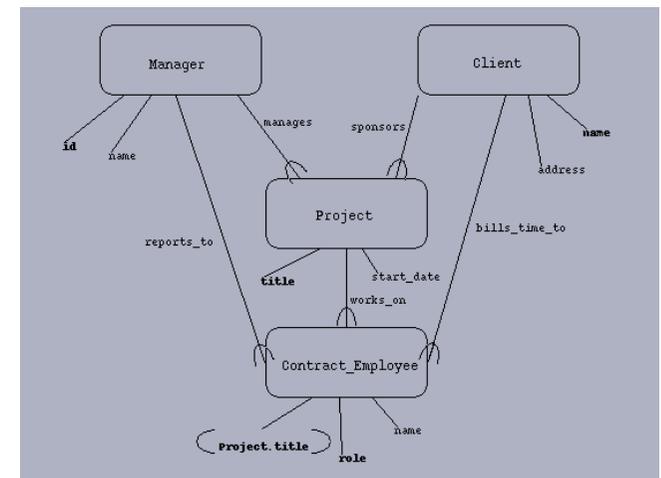
A System Description

A contracting company engages temporary (contract) employees to work solely on a single project, which is sponsored by a single client. The project is managed by one of the company's permanent staff. Every contract employee, who is identified within the company by the role they play in his or her project, is paid for in full by the client of the project on which (s)he is working, and the contract employee reports to a manager who is a member of the company's permanent staff. Contract employees, Managers and Clients all have names; Client names are unique. Projects are identified by their title, and Managers have a unique (company) identifier. Client addresses and the start date of each project are also recorded.

An initial ER model

This model contains two relationships - "bills time to" and "reports to" - which may be redundant.

It is essential that whether or not they mean the same as the composite paths - "works on - sponsors" and "works on - manages", respectively - is clarified with the client, rather than an assumption being made by the designer.



Path Composition

- Combining the Signatures of Relationships -

A relationship between two entities has four components: the degree (at each end), and the optionality (participation of each entity in relationship).

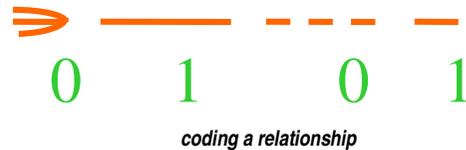
The signature of a relationship, in each direction, is either function or multifunction, and either total or partial.

The composition of relationship signatures shown here can be generalised for all relationship signatures.

Consider the four parts of a relationship, and code each part:

- 1 if single valued (function)
- 1 if mandatory (total)
- 0 otherwise

The overall signature of a composite relationship is the “AND” of the signature codes of the component relationships.



Researchers

David S. Bowers - Department of Computing, The Open University

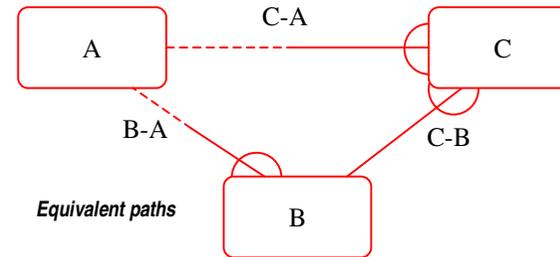
Recent Publications

Bowers, D.S (2000). 'Database Schema Improvement Techniques for CASE Tools' - UKAIS 2000.

Bowers, D.S (2002). 'Detection of Redundant Arcs in Conceptual Models' - International Workshop on Conceptual Model Quality, Turku, Finland

Collaborators: Alex Gray, Professor of Advanced Information Systems, Cardiff University

Syntactic Path Equivalence

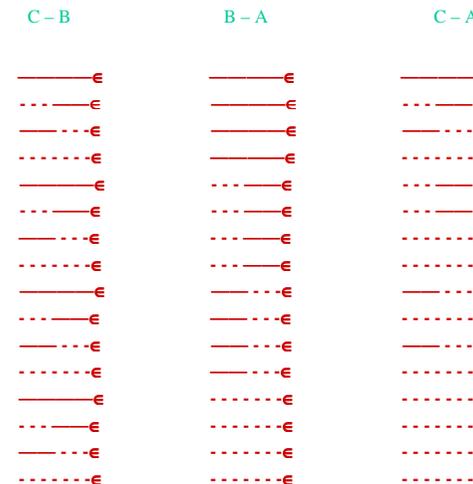


- Every “C” is related to one “B”;
- Every “B” is related to one “A”
- Hence, every “C” is related to exactly one “A” via one “B”.
- This signature is the same as that defined by relationship “C-A”
- Every “B” is related to one or more “C”
- Each “A” is related to zero or more “B”
- Some “A” are related to one or more “B” and hence, transitively, to one or more “C”
- ... which is the same signature, from “A” to “C”, defined by “C-A”.

The relationship “C-A” has the same signature, in each direction, as the path “C-B (B) B-A”.

Hence, it would be potentially redundant depending on whether or not it represented also the same semantics.

More General Cases

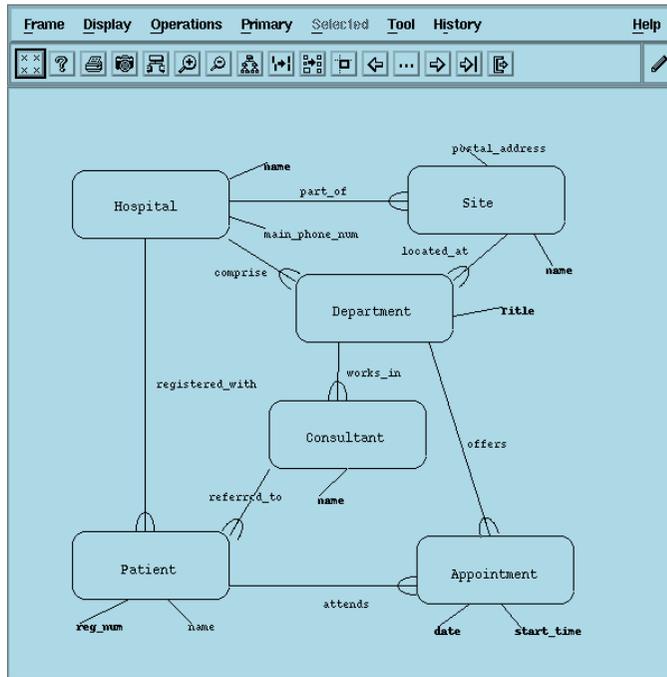


It is trivial to deduce these equivalent signatures for each of the 16 compositions of many:one relationships

The composition is recursive.

The ERD2 Tool in Action

An Initial Entity Relationship Model



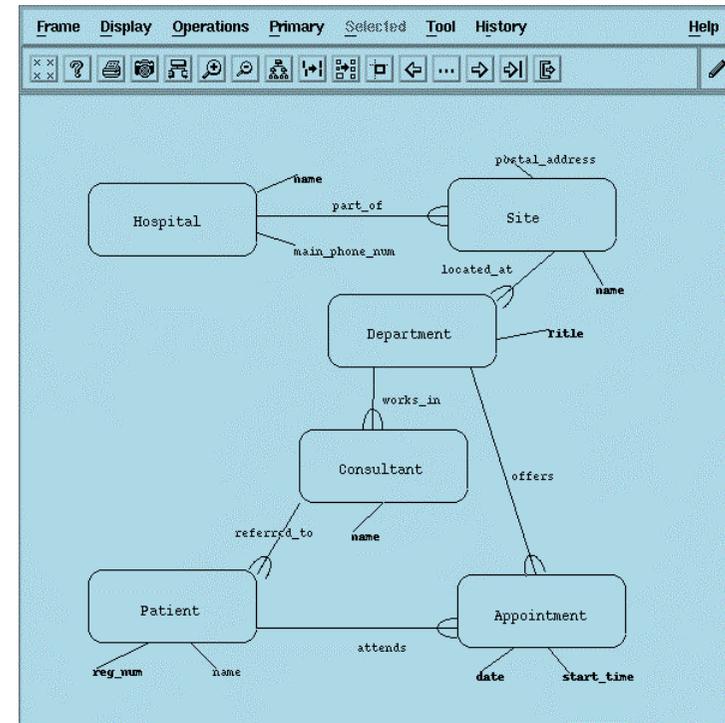
caption

Annotations

The Tool's Interactions with the Modeller

<p>Is "comprise" semantically equivalent to "part of", "located at"? [TRUE will delete "comprise"]</p> <p><input checked="" type="checkbox"/> TRUE <input type="checkbox"/> FALSE</p>	<p>Is "registered_with" semantically equivalent to "comprise", "works in", "referred to"? [TRUE will delete "registered_with"]</p> <p><input checked="" type="checkbox"/> TRUE <input type="checkbox"/> FALSE</p>	<p>Is "offers" semantically equivalent to "works in", "referred to", "attends"? [TRUE will delete "offers"]</p> <p><input checked="" type="checkbox"/> TRUE <input type="checkbox"/> FALSE</p>
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

The "corrected" Model



Researchers

David S. Bowers - Department of Computing, The Open University
 Kostas Vassiliadis, Andrea Levati, Mark Stuart - Final Year project students, University of Surrey

Recent Publications

Bowers, D.S (2000). 'Database Schema Improvement Techniques for CASE Tools' - UKAIS 2000.
 Bowers, D.S (2002). 'Detection of Redundant Arcs in Conceptual Models' - International Workshop on Conceptual Model Quality, Turku, Finland