

Database Reverse Engineering : A Form-Driven Approach

Nicodème Mfourga
University of Louvain, IAG-QANT, 1 place des doyens,
1348 Louvain-La-Neuve, Belgium, e-mail: mfourga@qant.ucl.ac.be.

Résumé

Ce papier présente une approche de rétro-conception des bases de données basée sur les formulaires utilisés pour la manipulation de ces bases de données.

1 Introduction

Assuming that many of the design elements (e.g., conceptual schema) which have led to a given database are lost when its evolution is needed, database reverse engineering is aimed at to produce these design elements which could reasonably be expected to have been produced as deliverables during the forward database design process, see e.g., [1]. Conceptual schemas provide a high level description of the information contents and can help users understand the database or start a new forward engineering process.

The main objective of our work is to develop an approach to reverse engineer database using principally data contained in database forms.

2 Database Reverse Engineering Technique Using Forms

By *form* we mean any structured collection of variables (form fields) which are appropriately formatted to communicate with database, especially for data retrieval and data display (paper forms and electronic forms (screen)). Such forms contain structural information and their fields have "natural links" which express domain semantics. The basic idea of our approach is that these links can be explored and exploited in order to infer database structural information. However, the relationships among fields on a display form may not be obvious in the layout. To make the structural relationships among components *explicit* and all components of the form individually "referenceable", we have defined a *form model* that will be generic enough that it permits to describe and abstract any database form to analyze it, while maintaining links with underlying database (see [2] for further details).

Our approach comprises two phases: *domain knowledge recovery* and *reconstruction of underlying database conceptual schema* that constitutes the final objective. To recover domain knowledge we have used a strategy borrowed from machine learning paradigm, *learning from examples*, that is, learning by generalizing specific facts or observations. The domain knowledge, e.g., data dependencies, is inferred from form instance examples by generalizing specific observations related to the data they display. Based on this knowledge, there are rules and heuristics that permit to extract an entity-relationship (sub)schema corresponding to a given form in a semi-automatic process aided by human (expert or analyst) activities (see [2]).

This approach can supplement existing database reverse engineering techniques where forms constitute important uses of the database.

3 References

- [1] Hainaut et al, *Contribution to a Theory of Database Reverse Engineering*, Working Conference on Reverse Engineering, Baltimore, Mai 1993.
- [2] Mfourga, N., *Using Forms in Database Reverse Engineering*, IAG-QANT, Univ. Catholique de Louvain, Belgium, Forthcoming.