
Strengths and Weaknesses of Database Models for Textual Documents

B.N. Rossiter† and M.A. Heather‡

† *Computing Laboratory, Newcastle University,
Newcastle upon Tyne, England NE1 7RU*

‡ *Sutherland Building, Newcastle Polytechnic,
Newcastle upon Tyne, England NE1 8ST*

ABSTRACT: User requirements in large and complex textbases are discussed in the light of current models. Examples applying relational and semantic models suggest criteria for a more fundamental approach involving the merger of object-oriented programming techniques with database methods in future complex object textbases.

KEYWORDS: document modelling, databases, complex objects.

1 Introduction

Little attention has been paid to text as structured data. Much of administrative data is in the form of textual strings but these tend to be treated as atomic entities independent of any relationship between words. Text retrieval and hypertext systems are based on physical divisions in documents and physical positions of words and rely on features like inversion, position operators and physical connections. By exploiting fully current technology such as multi-windowing and the emerging object-oriented programming, there have been significant advances in document manipulation in the provision of natural user interfaces [Pasquier-Boltuck et al 1988] and browsing systems [Brown 1988; Furuta and Stotts 1989]. There has been little regard for the very fine logical structure that lies beneath the physical form, even in recent data models for hypertext [Tomba 1989]. To handle large amounts of data of complex structure, more advanced file handling techniques will be required in the areas of full text information systems, electronic publishing, email, office automation, bulletin boards and conferencing.

Database technology needs to be extended from its present emphasis on simple objects to deal with complex objects such as text [Stonebraker et al 1987; Heather and Rossiter (in press)], CAD/CAM, CASE, knowledge