

bases and complex business information. In this paper, the work reported at WOODMAN'89 [Heather and Rossiter 1989] is refined with particular attention being paid to the semantics of class structures and symbolic keys. The aim is to show the potential of the developing object-oriented database technology for modelling textual documents.

2 Demands of Textual Applications on Filing Systems

As a first step in formulating a model, the demands made by textual applications on the technology of filing systems have been analyzed by drawing on applications at both Newcastle and elsewhere. The first two columns of figure 1 summarize the results which have been reported more fully elsewhere [Rossiter and Heather 1990]. This paper considers the implications of the requirements for database systems. Besides the obvious structural properties listed in section 1 of figure 1, many of the other needs also place demands on textual filing systems: context and proximity matching require fine index structures; thesauri introduce further types of data; referential transparency, to provide navigational facilities as in hypertext, requires links from one text to another ideally using symbolic identifiers; trails made while navigating text structures need to be recorded as fully-fledged data [Zellweger 1989; Sillitoe et al 1990]; and updating involves the major problems of version management and the modelling of dynamic behaviour such as the life cycle of a document.

Three inherent requirements for modelling text structures need to be satisfied for a successful database initiative: dynamic selection of unit size; representation of non-hierarchical text structures; and description of data in generalized and specialized forms.

Unlike informal systems which might store text as a continuous stream of characters, formal systems need an object unit. Traditionally a choice has been made governed by the storage capability of the system, by the human capacity for searching, retrieving and comprehending the information, and by the character of the document which will often be determined by traditional printing techniques. Thus, for example, the size of a legal statute is controlled by parliamentary business and other political factors. However, one unit alone is insufficient for all purposes. With text objects, it should be possible to retain the ability of natural language to keep the choice of unit dynamic and with the option of lazy evaluation to postpone any decision until the full circumstances and context of use are known.

Object-oriented
Semantic
E-R
Text
Relational
stand. extend.
Free Text

Fig
by Dat