

Chapter 9

The Separation of Data and Information in Database Systems under an Organisational Semiotics Framework

X. Wang, J. Feng

*Database Research Group
School of Computing
University of Paisley*

Abstract

Some significant problems in database systems research, such as query answering capability, connection traps, lossless transformation, and normalisation are difficult to explain, answer, solve, or explore further within the current context where data and information are fused, or even taken as the same thing. It would appear that a fresh and new perspective that distinguishes between data and information, and takes data as a type of sign, which bear (carry) information, is beneficial. The ideas of organisational semiotics (OS) enable this. We will discuss the reasons for separating data and information, and for investigating the relationship between the two. We will look at how and why the contemporary, seemingly muddled view, on the relationship between data and information might have hampered the progress on a number of database research issues. We conclude that looking at data and information within an OS framework sheds light on these issues and helps investigate them with a sound theoretical ground.

9.1 Introduction

The concepts of data and information are fundamental in database systems research. Although many researchers have been working on them, it would seem that unified definitions on data and information are unlikely to be reached in the near future.

Traditionally data is taken as raw material whilst information is generated from data after being summarised or analysed. One of the typical definitions is: "...data consists of the raw facts and figures that are processed into information. Information is summarised data or otherwise manipulated (processed) data" [13]. In Elmasri and Navathe [8], a database is explained as a collection of related data and is specified as facts that can be recorded and have implicit meaning. Checkland [3] proposes "information equals data plus meaning". "Information is data that has value. Informational value depends on context. Until it is placed in an appropriate context, data is not information, and once it ceases to be in that context it ceases to be information". [4]. Mingers [20] argues that "meaning is created from the information carried by signs. The consequences are that information is objective, but ultimately inaccessible to humans, who exclusively inhabit a world of meaning. Meaning is essentially intersubjective — that is, it is based on a shared consensual understanding."

Database systems are vehicles of storing and providing information [1]. Even through the above brief citations of relevant literatures, we find that most of the time information is fused with data. What you can see is what you can get in a database. One prevailing view is that data which has its meaning equals information in the context of database systems. And this view has been adopted as an implicit assumption in some database research, one of which is notably the Relative Information Capacity theory [14]. This theory reveals the fact that data instances of schemata are taken as "information" in databases and the "information capacity" of a schema is the capacity of accommodating instances into it. We argue that instances cannot be simply taken as the information in databases and that it would be a flawed theory if the confusion of data and information exists.

The rest of the chapter is structured as follows. We briefly state the theoretical foundations of this study and discuss the possibility of drawing from them to address the problem at hand in section 2. We give some analysis on the need of separating data from information in section 3. We enumerate some questions in database research, which have not been answered very well due to this confusion of data and information in section 4. The benefits of clarifying such basic and fundamental issues are also discussed in that section. And finally the conclusion of this work and our