Although the terms “data”, “knowledge”, and “information” are frequently used when referring to information systems, there is neither a clear distinction amongst them, nor a clear way to apply them in the development of database and knowledge-based systems. The objective of this paper is to analyze the current literature on data, knowledge and information, and strive to provide a set of guidelines for distinguishing amongst these terms. These guidelines could be useful to system users and developers to make efficient information resources management and utilization. In order to do so, semiotics is employed as a framework for analyzing existing interpretations of these terms. As a result of the analysis, five prominent features are identified which are then used to discuss data, knowledge, and information.

Current practices in the development of information systems suggest that the first, and most important step in the development process involves understanding the “domain” of the problem (This is referred to, for example, as the requirements specification and analysis phase in the database design process.). Fundamental to this understanding is the adoption of a common vocabulary that is meaningful to both the system’s users and its developers. It is, therefore, remarkable that, although the terms database system, knowledge-based system, and information system are used universally, the terms, “data”, “knowledge” and “information” themselves are rarely defined precisely and little agreement exists about the domain and scope encompassed by each [Wiederhold, 1986b; Bubenko and Orci, 1989]. Knowledge, for example, is often used as a synonym for either data or information. Information is usually used as the most general term and can mean either data or knowledge [Stamper, 1973]. If, however, data, knowledge, and information are different things with different properties, then one would expect different design issues to arise and different methods to be used in developing and implementing systems that must process them.

In this paper, the term “information system” refers to any system whose inputs and outputs consist of “signs” (Signs include numerical and alphabetical characters, words, sentences, messages of any length, and actions [Stamper, 1973]). Users of an information system provide input relating to the real world and/or use the outputs as a basis from which to make decisions. The outputs of an information system can also be used as inputs to other information systems or to the system itself. Figure 1 shows a generic information system model.

This research strives to provide an appropriate set of characteristics for distinguishing data, knowledge, and information, which could be useful to system users and developers. In order to carry out this task, existing interpretations of data, knowledge, and information are presented and examined using semiotics as a general framework for analysis.

Semiotics Framework

Data, knowledge, and information (as naively understood) require signs for their representation and processing by computers. Therefore, an analysis of the properties of signs
is a crucial first step in understanding the differences amongst these terms. Since *semiotics* is concerned with the properties of things in their capacity as signs [Morris, 1946], we have chosen to adopt it as a framework for our analysis. Semiotics is usually divided into three branches: pragmatics, semantic}, and syntactics [Morris, 1946]. Pragmatics deals with the origin, uses, and effects of signs within the behaviour in which they occur and, thus, explicitly includes the users of the signs. Semantics deals with the signification of signs; that is, the relationship between the signs and the objects to which the signs refer and ignores the user of the signs. Syntactics analyzes the relationships among signs without concern for the user or the signification of the signs [Carnap, 1942].

Figure 2 shows the scope of semiotics and illustrates the relationships of signs to users and real-world objects. Each branch of the semiotics framework is applied to existing interpretations of data, knowledge, and information below.

**Pragmatics**

Pragmatics deals with relationships between signs and behaviour [Stamper, 1973]. The way in which signs acquire their meaning is dependent on the behaviour of their users. Much of the research that focuses on the distinctions amongst data, knowledge, and information can be classified as having a pragmatic interpretation.

**Bell.** Bell [1979] distinguishes between information and knowledge based on how they should behave. Knowledge is referred to as an organized set of statements of facts or ideas which present a reasoned judgment or an experimental result. Information is defined as data processing (manipulation) in the broadest sense; that is, the result of storage, retrieval, and processing of data. This definition explicitly indicates the required operations for providing information.

**Newell.** Newell [1982] defines the following Principle of Rationality: *If an agent has knowledge that one of its actions will lead to one of its goals, then the agent will select that action.* This principle is clearly an extreme form of pragmatics, dealing only with the goal-seeking behaviour of an agent. In this context, *knowledge* is: “[w]hatever can be ascribed to an agent, such that its behavior can be computed according to the principle of rationality. Knowledge is to be characterized entirely functionally, in terms of what it does, not structurally, in terms of physical objects with particular properties and relations.” Thus, in specifying what knowledge is, it is important to understand why it is valuable and how it is used. The answers to these questions will serve as fundamental requirements for any system whose purpose is to store and process (representations of) knowledge. In addition, Newell distinguishes between: 1) knowledge itself, which is inherently abstract and exists functionally (the knowledge level); and 2) representations of knowledge which are models of knowledge (the symbol level). Furthermore, there are at least two ways in which such representations are useful: to the agent itself, as an aid in carrying out its goal-seeking behaviour, and to any observer (including the agent) who is trying to analyze the nature and extent of the knowledge possessed by the agent. Note that these two roles can have quite different implications for the nature of the representation scheme chosen. Newell also suggests that goals are a kind of knowledge having three constituents: knowledge of the desired state of affairs; knowledge that the state of affairs is desired; and knowledge of associated concerns, such as useful methods, prior attempts to attain the goals, etc.

**Bubenko and Orci.** Bubenko and Orč [1989] discuss three different views of knowledge: a database view, an Artificial Intelligence (AI) view, and an epistemic view. Their AI view is, in essence, the same as Newell’s. Bubenko and Orč [1989] present a classical definition of knowledge from philosophy, which they call the epistemic view: *C* knows *k* if (1) *k* is true, (2) *C* accepts *k*, and (3) *k* is evident for *C* where *k* is a statement. This implicitly defines the knowledge of the knower, *C*, presumably, as the set of all *k* that *C* knows, perhaps restricted to a specific topic. Clearly, the role of the knower is an inseparable part of the definition, and, therefore, this is a pragmatics view. Although, it is rare in practical matters that truth with certainty is possible, clause (1) makes the point that a statement, *k*, should be (approxi-

---

**Figure 1: A Generic Information System Model**

**Figure 2: The Semiotics Framework**
Related Content

Temporal Data Management and Processing with Column Oriented NoSQL Databases

Remarks on a Fuzzy Approach to Flexible Database Querying, Its Extension and Relation to Data Mining and Summarization
[www.igi-global.com/chapter/remarks-fuzzy-approach-flexible-database/52299?camid=4v1a](www.igi-global.com/chapter/remarks-fuzzy-approach-flexible-database/52299?camid=4v1a)

The Impact of Conflict Judgments between Developers and Testers in Software Development

Visual Query Languages, Representation Techniques, and Data Models
[www.igi-global.com/chapter/visual-query-languages-representation-techniques/28554?camid=4v1a](www.igi-global.com/chapter/visual-query-languages-representation-techniques/28554?camid=4v1a)