What is Semantics?

Dr. Tom Johnston
MindfulData.com
10/25/07

Definition.

Semantics:

1. In language, what gives words and phrases their meaning. Also, what gives sentences their meaning.

2. In data management, what gives data its meaning. More specifically, what gives instances of data schemas their meaning.

Commentary.

Semiotics.

Semiotics is the study of what gives any symbol or any object its meaning. For example, a child's sniffles have meaning for his mother; they mean that the child is probably coming down with a cold. Or consider the stylized icons outside toilet entrances in airports. By themselves, they are just geometrical shapes. It is our understanding of them as representing men or women that gives them their meaning. So, on the one hand, we have a pair of geometrical shapes. On the other hand, we have those shapes *as* indicating which toilets are for men and which are for women; and that is their meaning.

Semantics and Semiotics.

Semantics is a specialized branch of semiotics; it is the semiotics of language. So squiggles on a page or sound waves in the air correspond to the sniffles and shapes just mentioned. But it is almost impossible for we English speakers to look at those squiggles on a page (this page, and these squiggles, for example) or hear those sound waves, and not understand what they mean. Semantics is what gives them their meaning.

Semantics and Computer Scientists.

Computer scientists seem to assume that what “semantics” means is obvious. So they don’t define semantics, as it applies to data. Instead, they immediately use the word and starting discussing “semantic databases”, “semantic networks”, “semantic representations”, “truth-functional semantics”, “referential semantics”, etc, etc.

That's why we need to "roll our own" and provide our own definition of "semantics" as it applies to data.

The Semantics of Data. What Does Give Data Its Meaning?

There are two things that give data its meaning. One is the rules to which it conforms, and which are followed by those who create the data, and understood to have been followed by those who consume the data. The other is the interpretation given by the community of data creators and consumers to that data, and to the labels given to it.

Internal Contributions to Meaning.

Internally, the semantics of data is given by the rules to which instances of data conform. Those rules are:

1. Integrity rules.
    a. The entity integrity rule. Every entity instance must be uniquely identified by a subset of the data it contains.

    b. The referential integrity rule. Every required reference from one entity instance to another must be valid. If the reference is optional, the referencing link (foreign key) may be null.

    c. The domain integrity rule. Every value in a required attribute instance must be taken from a subset of values defined by the attributes data type and length.

2. Insert rules. Rules describing the necessary and sufficient conditions for creating a new entity instance.

3. Delete rules. Rules describing the necessary and sufficient conditions for deleting an entity instance.

4. State transition rules. Every update to a required attribute instance must result in a new value in that instance which is taken from the domain for that attribute; and in addition, if there are further rules constraining the update, they must be honored also. If an attribute is not required, it may be null, provided setting it to null does not violate state transition constraints.

5. Cardinality rules. Assuming that all relationships represented in a database are binary:
    a. The minimum cardinality rule for each of the related entities states whether the relationship is required or optional for that entity.

b. The maximum cardinality rule for each of the related entities states whether one or more than one instance of that entity may be related to one instance of the other entity.

External Contributions to Meaning.

Externally, the semantics of data is given by the interpretation we give to data and its labeling. Those interpretations are expressed in several different formats. They include:

1. Entity definitions. An entity definition should designate the domain of real world objects (physical or conceptual) which may be represented by an instance of an entity, and then also state the necessary and sufficient conditions for one member of that group of objects to be represented by an instance of an entity.

2. Attribute definitions. An attribute definition should state what property or relationship of the object instance is represented by that attribute.

3. Domain value definitions. A domain value definition should state, either individually or collectively, what the values of a domain represent.

4. Transformation definitions. For derived data, a transformation definition should describe the base values on which the derived value is based, and the computation or other rule used to create the derivation.

Combining the Internal and External Contributions to Meaning.

Both the internal and external components that supply semantics to data must be present in order for us to understand what data means.

1. Rules. Someone who knows something, and wants to put that knowledge into a database so others can access it, creates or modifies an entity instance in accordance with the internal rules stated above. Someone who accesses the data created by that person interprets that data in light of his belief that the person who created it followed the same rules that he (the consumer) believes were followed.

2. Labels. The labels – entity, attribute and relationship names – provide a broad context for understanding what data means. They do so because those who provided requirements for the database chose those labels based on the vocabulary they themselves use to talk about that subject matter.

3. Both. But without the precise rules governing the inserting, updating and deleting of entity instances, those labels would be just suggestions about what the data means. Knowing that those rules have been followed, those who retrieve that data then know a lot more about the entity instances in their result sets than just what

the English language labels mean. They know that those entity instances conform to those rules.

4. In Summary. Thus, semantics is provided for data by defining rules for creating instances of types of data, and following those rules, and also by labeling the data and defining the labels.
    a. When that is done, and we are looking at rows and columns of (usually) alphanumeric character strings in a query result set, we aren't just looking at squiggles on a screen or a page. We also aren't just looking at strings of letters and numerals. We are looking at data that *means* something to us.

    b. We now can have a deeper appreciation of what the distinction is between data and information. Information is data that is meaningful. ***Semantics is what turns data into information***.